Mechanistic Modelling for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations

Spann, Robert

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Robert Spann

Mechanistic Modelling for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations

PhD Thesis, July 2018

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Mechanistic Modelling for Risk-Based Monitoring of Lactic Acid Bacteria Cultivations

Robert Spann

Technical University of Denmark

Kgs. Lyngby, Denmark, 2018

Technical University of Denmark
Department of Chemical and Biochemical Engineering
Process and Systems Engineering Center (PROSYS)
Søltofts Plads Building 229
2800 Kgs. Lyngby
Denmark
Phone: (+45) 45 25 25 25
Email: dtu@dtu.dk
www.dtu.dk

Summary

There is increasing interest in process analytical technology (PAT) tools for on-line monitoring and control of bioprocesses. When modelling large-scale production processes, microbial processes, heterogeneous process conditions, and process uncertainties have to be considered. Heterogeneous process conditions arise because of insufficient mixing at a large scale, and they often have an impact on microbial activity and product yield as the cells have to adapt to changing process conditions when circulating through the bioreactor. Over the past years, computational fluid dynamics (CFD) and compartment models have been applied to account for fluid dynamics in bioreactors. They have been coupled with biokinetic models, and employed to study substrate and oxygen gradients. The focus of this work was on pH gradients in lactic acid bacteria cultivations as the pH is a critical process parameter for these cultivations. Process uncertainties are considered with a Monte Carlo simulation allowing risk-based decision making.

An aerotolerant *Streptococcus thermophilus* batch cultivation was utilized as a case study, where the cells are the target product as they are used in the dairy industry as starter cultures, e.g., for yogurt and cheese production. In this work, a mechanistic model was developed to describe the microbial cultivation, and it was applied together with a CFD and compartment model to account for mixing in the bioreactor. The model was applied to gain deeper process understanding, to test new process conditions, and as a soft sensor for risk-based monitoring of the cultivation.

The kinetic model consisted of a biokinetic model and chemical model. The biokinetic model described biomass growth, lactose (substrate) consumption, and lactic acid production (as a by-product) among others. The chemical

model was utilized to predict pH. A comprehensive parameter estimation was performed using lab-scale data in order to assess the identifiability, sensitivity, and uncertainty of model parameters. The model was comprehensively validated with independent batch and continuous cultivations exhibiting a relative mean error of less than 10 % for the biomass concentration.

The formation of pH gradients was investigated in a 700-L bioreactor. A one-phase CFD model and compartment model were developed and validated with tracer-pulse experiments to evaluate mixing time. A mixing time of 48 s was measured while the CFD and compartment models predicted 46 and 52 s at 240 rpm, respectively. The kinetic model was then coupled with both the CFD and compartment model, and a cultivation was simulated with both models. Besides an accurate prediction of the biokinetic state variables, both models predicted the pH gradients qualitatively with a deviation of less than 0.15 pH units. However, the CFD simulation took 4 days on 20 CPU cores, while the compartment model was solved within 2 s on an ordinary computer.

The compartment model was integrated in a probabilistic soft sensor for the monitoring of lactic acid bacteria cultivations. The soft sensor used the limited on-line measurements, namely the balance readout of the base (ammonia) addition and pH, to update the model parameters that were used as an input to the dynamic model. A Monte Carlo simulation of the model was performed to account for uncertainties in model parameters, initial process conditions, and on-line measurements. The soft sensor was validated with historical lab-scale and pilot-scale experiments. If the soft sensor was applied on-line during the production, it would equip plant operators with a tool to assess pH gradients on-line and to take risk-based actions. The probability of not achieving the target biomass yield was predicted based on the probabilistic model predictions. In the investigated cultivation, the risk was to lose 3.5 % of total production capacity. The operators could react accordingly, for example, by increasing the stirrer speed to reduce the pH gradient and achieve the desired production capacity.

Resumé

Der er en tiltagende interesse for værktøjer indenfor procesanalytisk teknologi (PAT) til on-line overvågning og kontrol af bioprocesser. Når storskala produktionsprocesser modelleres, skal forhold såsom mikrobielle processer, heterogenitet, og procesusikkerheder tages i betragtning. Heterogene procesforhold opstår i storskalaproduktion som følge af utilstrækkelig omrøring, som ofte har indflydelse på mikrobiel aktivitet og produktudbytte, eftersom mikroorganismerne løbende skal tilpasse sig de varierende procesbetingelser, som skyldes omrøringen i bioreaktoren. Computational fluid dynamics (CFD) og compartment modeller anvendes ofte til undersøgelse af gradienter i produktionstanke. Denne afhandling fokuserede på pH-gradienter i produktionsprocessen af mælkesyrebakterier, idet pH er en kritisk proces parameter i fermenteringerne. Som grundlag for risikobaseret beslutningstagning, vurderes processens usikkerheder ud fra Monte Carlo simuleringer.

Studiet var baseret på produktionen af mikroorganismer fra en aerotolerant batch-kultur af *Streptococcus thermophilus*, der i mejeriindustrien bruges som en start-kultur i for eksempel yoghurt og osteproduktion. En mekanistisk model blev udviklet og anvendt til beskrivelse af mikroorganismernes vækst. Der supplereres med CFD og compartment-modeller, som begge tager højde for opblandingsgraden i bioreaktoren. Modellens formål var at få en bedre procesforståelse, at teste nye procestilstande og at anvende den som soft sensor til risikobaseret overvågning af fermenteringen. Den mekanistiske model bestod både af en biokinetisk og en kemisk model. Den biokinetiske model beskrev blandt andet processer som biomassevækst, laktose (substrat) forbrug og mælkesyreproduktion (som biprodukt). Den kemiske model blev brugt til at beregne pH. En omfattende parameterestimering blev udført

til vurdering af statistisk identifikation, sensitivitet og usikkerhed i modelparametrene med udgangspunkt i laboratoriebaseret data. Uafhængigt data fra batch og kontinuerlige fermenteringer blev brugt til validering af modellen, som for biomassekoncentrationen viste en relativ gennemsnitsfejl på mindre end 10 %.

En ét fase CDF model samt en compartment model blev udviklet til undersøgelse af pH-gradienter. Begge modeller blev valideret i forhold til blandingstid bestemt i et sporstof-puls-eksperiment i en 700 L tank. Den eksperimentelle blandingstid blev målt til 48 sekunder, mens CFD modellen forudsagde 46 sekunder og compartment modellen 52 sekunder ved 240 omdrejninger pr. minut. Den kinetiske model blev koblet sammen med både CFD og compartment modellerne til kvalitativ forudsigelse af pH-gradienter med en afvigelse på mindre end 0.15 pH-enheder. CFD-simuleringen tog 4 dage på en 20-kerner CPU, mens en almindelig computer løste compartment modellen inden for 2 sekunder. Compartment modellen blev integreret i en probabilistisk soft sensor til overvågning af fermenteringsprocessen. Soft sensoren brugte tilgængelige on-line målinger af pH-værdier og doseret base (ammoniak) til opdatering af modelparametrene, som efterfølgende blev brugt som input til den dynamiske model. Yderligere blev Monte Carlo simuleringer foretaget for at tage højde for usikkerheder forbundet med modelparametre, processens startbetingelser og on-line målinger. Soft sensoren blev valideret i forhold til laboratorie og pilot-skala forsøg.

Implementering af on-line soft sensorer i processen ville kunne give operatørerne et værktøj til løbende at reagere på pH-gradienter ud fra risikobaserede handlinger. Sandsynligheden for at det ønskede procesudbytte ikke opnås, blev undersøgt ud fra en risikoanalyse. I den undersøgte proces blev en risiko på tab af 3.5 % af den totale produktionskapacitet estimeret. Procesoperatøren har så mulighed for at reagere i henhold hertil ved at reducere pH gradienten og dermed sikre den ønskede produktionskapacitet, eksempelvis gennem en forøgelse af omrøringshastigheden.

Preface

This Ph.D. project was conducted at the Department of Chemical and Biochemical Engineering (Kemiteknik) at the Technical University of Denmark (DTU) during the period from August 2015 until July 2018. This project was part of the European Union's Horizon 2020 Marie Skłodowska-Curie Actions Biorapid project that aimed to develop modelling and monitoring methods for the process development of bioactive molecules. The main idea of this Ph.D. project was therefore to develop modelling tools, to validate them with lactic acid bacteria cultivations as a proof of concept, and to contribute to research in the field of bioprocess development.

This project was conducted in collaboration with Chr. Hansen A/S and the Chair of Bioprocess Engineering, Department of Biotechnology at the Technische Universität Berlin (TU Berlin). The experiments of this thesis were carried out there during secondments, partially together with Klaus Pellicer Alborch (TU Berlin). The remaining work was performed at DTU.

Associate Professor Gürkan Sin (DTU) was my main supervisor with Professor Krist V. Gernaey (DTU), Anna Eliasson Lantz (DTU), Christophe Roca (Chr. Hansen), and David Kold (Chr. Hansen) as co-supervisors.

This thesis consists of two parts:

- Part I summarizes the highlights of the results;
- Part II consists of the journal articles and peer-reviewed conference proceedings that were written during my Ph.D. study.

Kongens Lyngby, July 2018
Robert Spann

Acknowledgments

Many people have supported me during the past three years, and I have greatly enjoyed working with them. I would like to thank my supervisors, Gürkan Sin, Krist V. Gernaey, Anna Eliasson Lantz, Christophe Roca, and David Kold for the fruitful discussions and their help during my studies. They have always provided constructive feedback, and I enjoyed talking about topics from different angles with them.

From Chr. Hansen, I would also like to thank the people who supported and helped me during my experiments. I very much appreciate the industrial insight that they provided as this was very important for this project and encouraged me to continue research on this topic.

I have genuinely enjoyed working at the Technische Universität Berlin during my secondment. It was a pleasure to work with Klaus Pellicer Alborch. I would like to thank, in particular, Peter Neubauer and Stefan Junne who already greatly supported me during my earlier studies and largely contributed to my deep interest in biotechnology.

At the Technical University of Denmark, I would like to thank my friends and colleagues, who made every day a special day. I truly enjoyed coming to DTU every day. With this, I wish to thank the students, Pinxiang Han, Jens Glibstrup, and Peter Nielsen, who contributed to this project.

I would especially like to thank my wife, Lisa, and my other family members and friends for their support and encouragement over the past years.

This project received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No 643056 (Biorapid project).

Table of Contents

Introduction

The aim of this project is to develop and validate a comprehensive mechanistic modelling framework to monitor lactic acid bacteria cultivations under consideration of pH gradients and uncertainties. An aerotolerant *Streptococcus thermophilus* batch cultivation is employed as a case study. The target cultivation products are the cells which are then used in the dairy industry, e.g., as starter cultures for yogurt and cheese production.

The focus of this project is to assess pH gradients and to take uncertainties in the dynamic model simulation into account (Fig. A). pH gradients occur because of insufficient mixing in large-scale bioreactors and they are regarded as one of the key factors affecting the biomass productivity. CFD and compartment models are applied to model the concentration heterogeneities

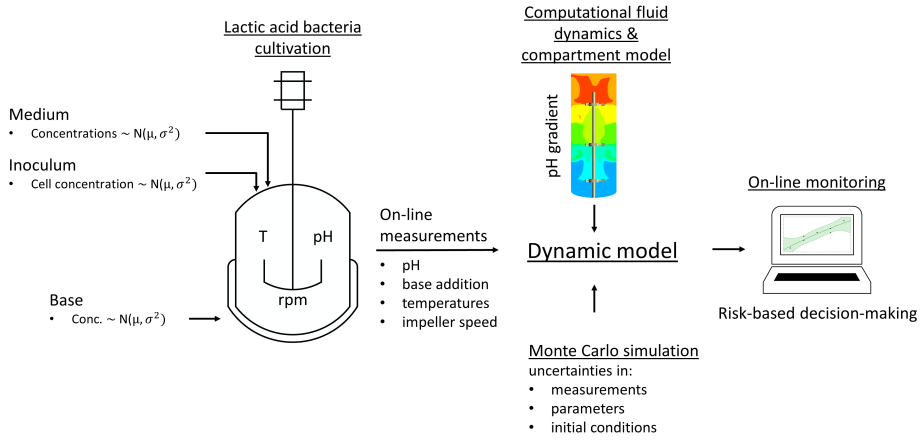


Figure A: Overview of this project. A PAT tool for risk-based monitoring of lactic acid bacteria cultivations.

in the bioreactor. Uncertainties arise naturally in the process as, for example, substrate concentration varies slightly from batch to batch. In addition, model parameters are subject to data quality in the parameter-estimation step. A Monte Carlo simulation is applied to account for these uncertainties in the model predictions. The final outcome of this project is a soft sensor for risk-based monitoring of the cultivation. It uses on-line measurements to update the dynamic model, CFD model and compartment model to consider pH gradients, Monte Carlo simulation to yield robust model predictions, and provides a PAT tool for risk-based decision making.

Structure of the Thesis

The thesis is structured in two parts. **Part I** consists of five chapters and the overall conclusions. First, the current state of the art in the field is described and then the main research results are explained. **Part II** contains the manuscripts, submitted and accepted articles, and peer-reviewed conference proceedings. They embody the details of the methodologies and provide further details on the results obtained.

Chapter 1: Literature Review

The literature review provides the background of the thesis. Heterogeneities that arise in large-scale bioprocesses are addressed together with appropriate modelling tools, namely CFD and compartment models. Furthermore, the state of the art of traditional one-compartment models is described, including the Monte Carlo procedure, in order to account for uncertainties.

Chapter 2: Model Development and Validation

Chapter 2 surrounds the mechanistic model that lays the foundation of this thesis. Firstly, a biokinetic model based on first principles is developed that describes the biomass growth of *S. thermophilus*, carbon source consumption, and lactic acid production as a by-product, among others. It is combined with a chemical model that predicts the pH of the cultivation broth. Secondly, parameter estimation, including identifiability, sensitivity, and uncertainty analysis, is performed to fit the biokinetic model parameters. The

model is then validated with independent data from batch and continuous cultivations.

Chapter 3: Probabilistic On-line Monitoring Using a Soft Sensor

The mechanistic model is applied in a soft sensor to monitor lactic acid bacteria cultivation. The soft sensor predicts both the current state of the cultivation as well as that in the future. A Monte Carlo simulation is included in the soft sensor to account for uncertainties in the model parameters, initial process conditions, and measurements. The soft sensor framework is developed and tested with historical data sets of lab-scale cultivations. So far, homogeneous process conditions are assumed.

Chapter 4: Prediction of pH Gradients in Lactic Acid Bacteria Cultivations

pH gradients in a 700-L bioreactor are investigated during a cultivation. On the one hand, multi-position pH measurements are performed in the bioreactor during the cultivation. On the other hand, the pH gradients are predicted in simulations. Firstly, a CFD model is applied to consider fluid dynamics and coupled with the kinetic model to simulate a cultivation. Mixing experiments and simulations are performed to validate the model. Secondly, a compartment model is designed starting from a CFD model. It is applied to the cultivation, and the results are compared with the experiments and CFD simulation. Besides the prediction accuracy of the pH gradients, the focus is on computational speed because this is an essential aspect in view of future on-line model applications as intended in this work.

Chapter 5: On-line Risk Quantification

The compartment model is employed in the soft sensor, and validated with a 700-L cultivation. The pH gradient in the bioreactor and the other state variables are predicted. In addition, the probability of not achieving the target biomass yield and risk of product loss are quantified in the soft sensor using the probabilistic model predictions. With this application, both the pH

gradient and the risk can be critically assessed, and they allow risk-based decision making by the plant operators.

List of the Main Contributions

The thesis is based on research results that have been included in a book chapter, journal articles and peer-reviewed conference proceedings. The articles and manuscripts could be found in **Part II**.

(A) Book chapter:
Spann, R., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). Modelling for process risk assessment in industrial bioprocesses. In Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. **Published**
https://doi.org/10.1016/B978-0-12-409547-2.14356-2

(B) Journal article:
Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. Biochemical Engineering Journal, 135, 49–60. **Published**
https://doi.org/10.1016/j.bej.2018.03.016

(C) Peer-reviewed conference proceedings:
Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2017). A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model. In A. Espuña, M. Graelss, & L. Puigjaner (Eds.), Computer Aided Chemical Engineering, Elsevier, 40, 2221-2226. **Published**
http://dx.doi.org/10.1016/B978-0-444-63965-3.50372-X

(D) Peer-reviewed conference proceedings:
Spann, R., Eliasson Lantz, A., Roca, C., Gernaey, K. V., & Sin, G. (2018). Model-based process development for a continuous lactic acid bacteria fermentation. In F. Anton, J. J. Klemes, S. Radl, P. S. Varbanov, & T. Wallek (Eds.), Computer Aided Chemical Engineering, Elsevier, 43, 1601-1606. **Published**
https://doi.org/10.1016/B978-0-444-64235-6.50279-5

(E) Journal article:
Spann, R., Glibstrup, J., Pellicer Alborch, K., Junne, S., Neubauer, P., Roca, C., Kold, D., Eliasson Lantz, A., Sin, G., Gernaey, K. V., & Krühne, U. CFD predicted pH gradients in lactic acid bacteria cultivations. Biotechnology and Bioengineering. **In preparation**

(F) Journal article:
Spann, R., Roca, C., Gernaey, K. V., & Sin, G. A Validated CFD-based Compartment Model to Assess pH Gradients in Lactic Acid Bacteria Cultivations. AIChE Journal. **Submitted**

(G) Journal article:
Spann, R., Gernaey, K. V., & Sin, G. On-line Process Risk Assessment of a 700 L Lactic Acid Bacteria Cultivation. Frontiers in Bioengineering and Biotechnology. **Submitted**

Part I

Chapter 1

Literature Review

The literature review is reprinted from the following book chapter with minor changes to improve the readability of this thesis and with the addition of the subsection, Lactic Acid Bacteria Models:

Paper A: Spann, R., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). Modelling for process risk assessment in industrial bioprocesses. In Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Elsevier.

1.1 Simulation of Heterogeneities at a Large Scale

"Mechanistic models are applied to various PAT systems but homogeneous culture conditions are assumed typically, even though heterogeneities occur at a large scale. In industrial large-scale bioprocesses, culture parameters, like the substrate concentration, oxygen concentration, pH, and temperature, are not homogeneously distributed, and gradients of them exist instead within the bioreactor (Lara et al., 2006). This is because of a lower mixing capability of large-scale bioreactors, which are on the scale of 10 to larger than 300 m^3 compared to small laboratory-scale bioreactors, which are in the litre range. Cells circulating within large-scale bioreactors are consequently exposed to continuously changing conditions, and they need to adapt to these conditions constantly, which affects their metabolic activity. As a result, biomass and product yields are often lower at large scales than at the laboratory scale (Bylund et al., 1998; Zou et al., 2012; Enfors et al., 2001; de Jonge et al.,

2011; Xu et al., 1999; George et al., 1998). It is also possible that strains that showed a good productivity at the laboratory scale do not grow at all at a large scale. The challenge is therefore to predict the behaviour of the microorganisms – mainly their productivity – at a large scale without the need for expensive experiments at the large scale to support optimal bioprocess development and control.

Gradients at a Large Scale

Gradients occur at a large scale because certain components, like substrate, oxygen, and base or acid, are added at one (or sometimes several) positions locally to the cultivation, and it takes then several seconds up to minutes to distribute them homogeneously throughout the bioreactor (Delvigne et al., 2006). This leads to a heterogeneous distribution of culture parameters, such as substrate, oxygen, carbon dioxide concentration, pH, and temperature (Fig. 1.1). Moreover, there exist pressure gradients in bioreactors owing to the height of the bioreactor (Neubauer and Junne, 2016), and cells are exposed to changing flow conditions (Nienow, 2006, 2014). The effect of gradients in large-scale cultivations is often a reduced biomass yield and productivity compared to homogeneous lab-scale cultivations. Cells that travel through a large-scale bioreactor have to deal with changing conditions uninterruptedly, which induces stress upon microorganisms. All microorganisms respond in a different way to the various oscillating conditions. It is therefore necessary to study the strain-specific response in so-called scale-down experiments, where large-scale conditions are mimicked in lab-scale experiments so that the microbial response can be investigated (Neubauer and Junne, 2010; Wang et al., 2014). The effects of substrate, oxygen, and pH gradients have been studied in detail for many organisms and certain examples will be highlighted in the following paragraphs. The objective of these studies is to investigate: i) how microorganisms respond to shifts in different conditions (or combinations of them); ii) how fast the organisms react to changes at the -omics level (genomic, transcriptomic, proteomic, and metabolomic); and iii) what the long-term effects of oscillating conditions are.

Substrate gradients play a major role in fed-batch cultivations, where

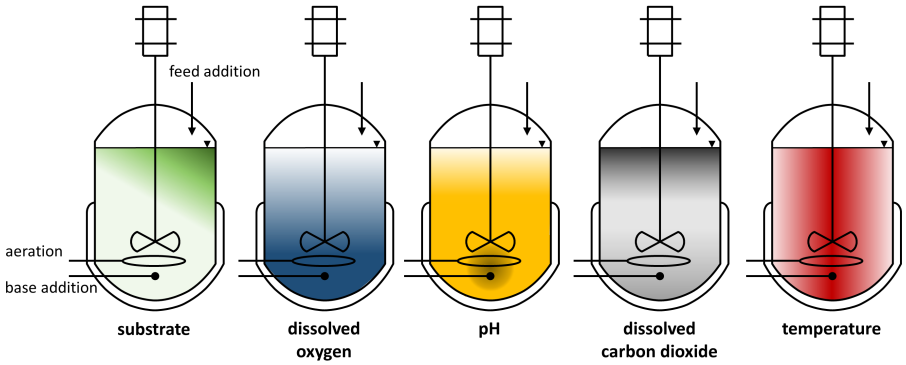


Figure 1.1: Gradients in large-scale cultivations. The formation of gradients for a fed-batch bioreactor configuration with the substrate-feed addition at the top of the bioreactor, and the oxygen sparging and base addition at the bottom of the bioreactor. Dark colour: high concentration/value; bright colour: low concentration/value. Reprinted from **Paper A**.

the substrate is usually fed at the top of the bioreactor (Fig. 1.1, substrate). High substrate concentrations of up to 2 g L^{-1} have been reported near the feed-addition point (Bylund et al., 1998). In the feeding zone, the cells take up much substrate according to their maximum substrate uptake capacity. This leads to a fast depletion of oxygen there when high cell densities are reached as high substrate turnover demands a high respiratory capacity. A further consequence is the formation of glycolytic overflow metabolites, such as acetate, lactate, formate, and succinate in *E. coli* cultivations (Sunya et al., 2013; Xu et al., 1999) or ethanol in *S. cerevisiae* cultivations (George et al., 1993) as the entire carbon cannot be metabolized by respiratory metabolism. However, these metabolites are then typically taken up again and re-metabolized in aerobic substrate-limited or -depleted zones, further away from the feed addition. Nevertheless, this generation and uptake of side products is a futile cycle, and can lead to a decreased biomass or product yield in the end as these detours are less energetically efficient. Recent analyses also investigated the response on the transcriptomic and proteomic level of *E. coli* and *C. glutamicum* (Lemoine et al., 2016; Löffler et al., 2017; Simen et al., 2017; Sunya et al., 2013), which allow for a detailed understanding of the microorganisms. This might allow genetically engineering microorganisms

that are less prone to stress, as shown in by Michalowski et al. (2017).

Oxygen gradients are particularly crucial when the cultivation reaches high cell densities and the available oxygen is consumed rapidly (Fig. 1.1, dissolved oxygen). Oxygen is added to the culture medium at the bottom of the bioreactor as gas (air or pure oxygen). Based on low solubility of oxygen in water, the available oxygen is limited, especially further away from the gas inlet, and in the feeding zone, where the substrate and oxygen uptake rates are high. Oscillating oxygen conditions have demonstrably reduced recombinant protein production in *E. coli* (Sandoval-Basurto et al., 2005) and expression of non-proteinogenic branched-chain amino acids (Soini et al., 2008), which could result in a misincorporation (an erroneous incorporation) into recombinant proteins. The lipase-producing strain, *Yarrowia lipolityca*, for instance, shows a reduction of lipase gene expression (Kar et al., 2008). Other strains, such as *Corynebacterium glutamicum*, seem to be robust with respect to substrate and oxygen gradients (Käß et al., 2014; Limberg et al., 2017).

pH gradients exist in bioreactors because the pH of the cultivation broth is controlled by adding base or acid as needed by a pH control setup (Fig. 1.1, pH). In aerated cultivations, the acid/base is often added together with the incoming aeration gas that ensures a very fast distribution of the acid/base. In non-aerated cultivations, the acid/base is usually added to the bioreactor as liquid, either at the top or at another position, e.g., close to the impeller blades. The acid/base addition leads to a zone of unfavourable pH conditions in close vicinity of the dosage point. Cells that pass this zone are prone to cell damage. pH shifts of almost 1 pH unit have been measured close to the alkali addition point in an 8-m^3 reactor for mammalian cell cultures (Langheinrich and Nienow, 1999), but even higher shifts might be expected. It was shown for *E. coli*, *B. subtilis*, and mammalian cells that oscillating pH conditions can impact biomass growth, the metabolome, the transcriptome, and cell viability (Amanullah et al., 2001; Cortés et al., 2016; Osman et al., 2002). Extracellular pH has a direct effect on cell physiology because it affects intracellular pH, which is crucial for enzymatic activity and controlled by proton pumps (Hansen et al., 2016).

CO_2 gradients probably have a major effect on the productivity of large-scale cultivations, too, but have been rarely investigated thus far (Fig. 1.1, dissolved carbon dioxide) (Baez et al., 2011; Blombach et al., 2013; Buchholz et al., 2014).

The effects of heterogeneities on microbial activity and productivity are a combination of the fluid dynamics in the bioreactor and microbial metabolism. Computational tools that describe the fluid dynamics are therefore coupled with biokinetic modelling in order to support the investigation of the effect of gradients at large scales.

Computational Fluid Dynamics (CFD) Models

CFD simulations can predict among other things the motion of the fluid in bioreactors and are therefore applied, e.g., for the prediction of fluid dynamics and mixing time when different impeller designs are tested for a bioreactor (Yang et al., 2012; Zou et al., 2012). Furthermore, CFD simulations can be combined with biokinetic models in order to investigate the effects attributed to gradients.

The fluid dynamics are numerically solved for the liquid volume in a CFD simulation. The fundamental mathematical equations are based on the conservation of mass, momentum, and energy. The most commonly used mathematical formulation is that of the Navier-Stokes equations. Further equations describing other phenomena, such as turbulence or eddy formation, are applied and solved depending on the goal of the simulation, as well. The liquid volume is divided into many (up to a few million) small elements for the simulation, and the fluid dynamic equations are solved for each element.

CFD is applied for biotechnological cultivations because many key issues in cultivations are dependent on the flow: mass transfer (e.g., mixing of feed streams; gas-liquid transfer), shear rates, and transport of microorganisms through the bioreactor (Delvigne et al., 2017; Schmalzriedt et al., 2003). To this end, the biological kinetic equations are solved together with fluid flow within the CFD simulations. The simulation then describes concentration gradients (e.g., substrate, dissolved oxygen), gradients of physical parameters (e.g., pH, mass transfer coefficients, gas hold-up), and temporal and

spatial performance of the microorganisms (e.g., substrate uptake, product formation, by-product formation, growth). The purpose of these studies is usually to predict the oscillating culture conditions at a large scale and design controlled scale-down experiments at the lab scale that mimic large-scale conditions as closely as possible. Future interest lies in the prediction of culture performance at a large scale using CFD simulations.

As a first step, the fluid dynamics of the bioreactor, i.e., the velocity profile of the bioreactor, is simulated. Second, the mixing time is often simulated. The mixing time is the time that a tracer that was pulsed into the bioreactor needs to reach a homogeneous distribution (i.e., to achieve complete mixing) in the bioreactor. The predicted mixing time is used to validate the CFD simulation as the simulated and experimental mixing times can be compared. Once the CFD simulation is validated, it can be applied to simulate microbial cultivations. The CFD simulation is then coupled with further gas-liquid transfer processes and biokinetic models to study bioprocess yields and performance in bioreactors.

There are two main approaches to model microorganisms in the CFD simulation, which are applied depending on the purpose of the simulation – the Euler-Euler and the Euler-Lagrange approach. With the Euler-Euler approach, the microorganisms are treated as a continuum, i.e., all cells are treated in the same way as concentrations of a dissolved component (Bannari et al., 2012; Elqotbi et al., 2013). With the Euler-Lagrange approach, the fluid is treated as a continuum, but the microorganisms are seen from the microbial point of view, i.e., individual cells travelling through the bioreactor are tracked (Haringa et al., 2016; Lapin et al., 2006; Morchain et al., 2013). The advantage of the Euler-Euler approach is that it is computationally less demanding than the Euler-Lagrange approach. Nevertheless, this leads to a loss of the individual history of the cells (Lapin et al., 2004). This could be of interest when one wants to investigate, e.g., the effect of subsequent, unfavourable conditions or culture conditions experienced by the microorganisms at a large scale over time, or so-called lifelines of the cells (Haringa et al., 2016; Kuschel et al., 2017). Scale-down experiments that resemble large-scale conditions could be designed based on the CFD simulation. Haringa et al. (2017) recently proposed a three-compartment lab-scale setup to mimic the

conditions of a 22-m^3 *S. cerevisiae* cultivation, which needs to be validated as a next step. CFD-based scale-down experiments will improve the quality of scale-down studies. CFD models are furthermore applied, for example, to enhance the treatment efficiency of wastewater plants (Karpinska and Bridgeman, 2016; Samstag et al., 2016).

Even though the CFD simulation of the fluid dynamics together with biokinetic models is a powerful tool that provides accurate predictions, its application is limited to off-line studies. This is because the CFD simulation, with its many mesh elements, is computationally too demanding, especially when large-scale bioreactors must be modelled. For on-line applications, such as monitoring and control, CFD simulations cannot be executed fast enough (e.g., every couple of minutes), and therefore compartment models are preferred.

Compartment Models

Compartment models (CMs, also known as block models or network of zones) divide the liquid domain into a limited number of coarse elements, which is much lower than the number of elements used in CFD simulations, and demands therefore less computational time compared to the CFD model. The properties (e.g., temperature, pH, concentrations) are assumed homogeneous in each compartment, and the Navier-Stokes equation is not included. Additional models, such as chemical, physical, biokinetic, or population balance models, are solved only with an affordable simulation time. The compartments are interconnected with defined flow rates and the interface area between each adjacent compartment.

A consistent framework to build CMs is based on an initial CFD simulation that provides a validated steady-state solution of the fluid dynamics (Moullec et al., 2010) (Fig. 1.2). The objective is to define the compartments based on the CFD results, preferably automatically according to a specific algorithm (Bezzo and Macchietto, 2004; Delafosse et al., 2010; Rigopoulos and Jones, 2003). Using the simulated velocity profile, compartments could be automatically defined, e.g. by identifying dead regions or recirculation regions (Bezzo and Macchietto, 2004). However, this is not a general solution

as adaptations for different kinds of bioreactors (e.g., plug-flow reactors) or impeller configurations are needed. Moreover, the derived compartments do not necessarily represent, for example, a feeding or acid addition zone, which might be crucial for several applications. Bezzo and Macchietto (2004) have therefore proposed a hybrid algorithm that takes also other properties, e.g., concentrations, into account for the identification of compartments. This methodology has the disadvantage that also a CFD simulation, including bio-kinetics, has to be performed once, in order to design the compartments. Nevertheless, the developed CFD-based CM can be validated in this way to replace the CFD simulation for accurate on-line applications. Manual definitions of compartments are also applied (Bashiri et al., 2016; Nauha et al., 2018). Delafosse et al. (2014) implemented homogeneously distributed compartments based on their coordinates, though they needed more than 12000 compartments in order to match the CFD-predicted and experimental mixing time of a 16 L bioreactor, which used more than 10^6 mesh elements for the CFD model. Such approaches will not be feasible for larger bioreactors or on-line applications combined with Monte-Carlo simulations (see probabilistic model predictions).

Once the compartments are defined, the volume flows (liquid and gaseous) between the compartments are taken from the CFD simulation. Further properties, such as volumes, pressure, shear rates, viscosities, etc. are also derived from the initial CFD simulation and used in the CM. Then, the CM needs to be validated, e.g., with mixing experiments or the already-validated CFD simulation (Delafosse et al., 2015). Finally, biological, physical, and chemical kinetics can be incorporated and then solved in all compartments together with the transfer processes between the compartments.

CMs have been applied to characterize the culture conditions in bioreactors (Bezzo et al., 2003; Guha et al., 2006; Moullec et al., 2010; Nauha et al., 2018; Rigopoulos and Jones, 2003; Wells and Ray, 2005; Zhao et al., 2017). Moreover, entire bioprocesses have been simulated. Glucose and acetate gradients could be predicted in a large-scale *E. coli* cultivation, and resembled the measured tendencies (Vrábel et al., 2001). An antibiotic-producing *Streptomyces fradiae* cultivation in a 3-m^3 bioreactor was modelled and could be validated with dissolved oxygen measurements (Vlaev et al., 2000). The

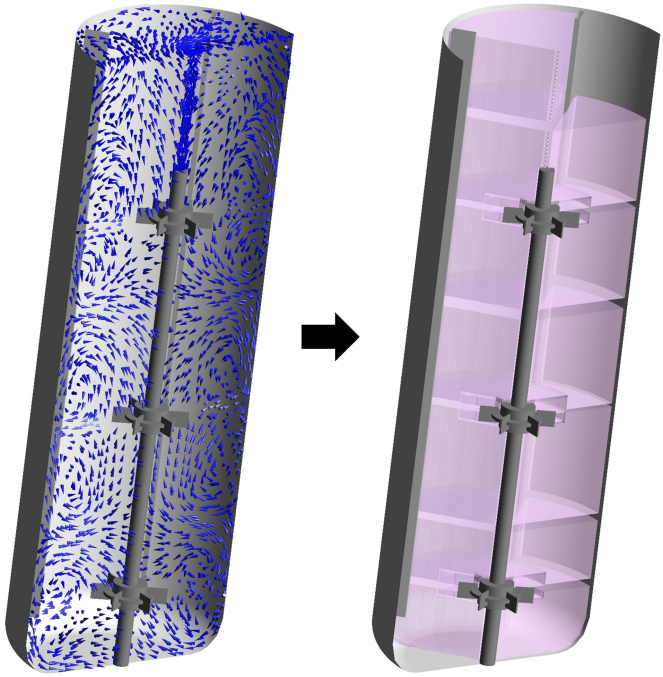


Figure 1.2: Example of a CFD-based compartment model. The steady-state velocity profile (left) can be used to define the compartments (right) for a compartment simulation. The velocity direction and magnitude (left, blue arrows) and an example of the compartments (right, pink boxes) are shown for a bioreactor with three Rushton turbines and four baffles. Reprinted from **Paper A**.

same framework has been applied to predict the dissolved oxygen tension in a 31-m^3 stirred-tank bioreactor for the production of another antibiotic, and for a 236-m^3 bubble column for citric acid production, however without validation (Zahradník et al., 2001). Different gradients (substrate, oxygen, etc.) were predicted for wastewater-treatment bioreactors (Alvarado et al., 2012; Moullec et al., 2010; Rehman et al., 2017). These studies included the biokinetics of the ASM1 model (Henze et al., 2000) to describe microbial conversions in the system in very large aeration tanks of several 1000 m^3. For a 30-m^3 *E. coli* cultivation, gradients for the glucose concentration, growth rate, and others were predicted, whereas a population balance was applied to account for cell heterogeneities (Pigou and Morchain, 2015; Pigou et al., 2017)." (**Paper A**)

1.2 Mechanistic Modelling of Lactic Acid Bacteria

"A mechanistic model describes the behaviour of a system with mathematical equations. Mechanistic models are applied to develop, optimize, control, etc. different bioprocesses. They are based on prior knowledge of the phenomena of the system, which are in particular the elemental mass, energy, and momentum balances of bioprocesses (Esener et al., 1983; Roels, 1981). Mechanistic models are therefore also called first principles, fundamental, or white-box models. Generally speaking, ordinary differential equations (ODEs) or partial differential equations (PDEs) describe the system if the model outputs change dynamically in time, or in time and space, respectively. The differential equations for all system components cover typically biological, chemical, and physical mechanisms, such as microbial growth, pH calculation, and aeration, respectively (see subsection: Biological, Chemical, and Physical Model Expressions for Cultivations). The main components are, for instance, the substrate (the carbon source), biomass, oxygen, reactor volume, and gas flows in an aerobic cultivation.

The main advantage of mechanistic models over data-driven models is that they have a large extrapolation capability as they are based on first principles and are not limited to the conditions that were employed to calibrate the model (Mears et al., 2017b). Mechanistic models have therefore a large application range (see section: Applications of Process Models) and can be utilized to test scenarios even if no data for these conditions are available. Another advantage of mechanistic models is that the model parameters have a physiological meaning, e.g., the specific maximum growth rate of the cells; hence, the user of the model can directly understand them. Furthermore, such models have a general and flexible structure that allows knowledge transfer of process equations and parameters both in industrial and academic environments from one process model to another. However, disadvantages of mechanistic models are that they require a relatively long development time, and that significant process insight is needed to obtain and validate the fundamental process equations.

In the end, mechanistic models for the bio-based production industry are often a combination of mechanistic and empirical equations based on

the complexity of the system. The latter is used especially for situations where certain phenomena are not understood to the necessary level of detail. There are also several other types of models for bioprocesses that are out of scope of this contribution, including data-driven models (which include, for instance, artificial neural networks and chemometric methods, such as principal component regression) and hybrid models (a combination of mechanistic and data-driven modelling techniques) (Solle et al., 2017; von Stosch et al., 2014).

The biotechnological industry increasingly applies mechanistic models because it has realized their significance. Mechanistic models are, for example, advantageous for predicting system behaviour after a change, while statistical models based on QbD sometimes fail (Process Development Forum, 2014). Today data-driven approaches are preferred in industry because of the cost-benefit analysis, which might be based on the established regulatory requirements for the approval of new processes. However, mechanistic models can support the PAT framework to better understand the effect of process changes under a wider range of conditions. Mechanistic models can simulate different initial conditions as well as process disturbances thanks to their extrapolation capabilities. They use the critical process parameters (CPPs) as inputs to predict critical quality attributes (CQAs) of, e.g., the desired product.

Once the model structure is established, the model parameters have to be determined and the reliability of the model must be assessed. It is important to prove the reliability of the model in order to be able to rely on model-based application, and to be able to defend, for example, the design space against the FDA. Several methods and tools exist to assess the credibility of the model, including identifiability, uncertainty, and sensitivity analysis (Chis et al., 2011; Sin and Gernaey, 2016).

Biological, Chemical, and Physical Model Expressions for Cultivations

The differential model equations typically cover biological, chemical, and physical processes. Biological processes are often modelled with a

macroscopic view, describing the microbial cell as a whole (unstructured model), especially when the model is applied for large-scale process design and optimization (Gernaey et al., 2010; Kroll et al., 2017). In structured models, at least one intracellular metabolic component, such as metabolites, ATP, and NADH, are described. This is of particular interest for metabolic engineering and microbial cell development in system biology (Almquist et al., 2014; Campbell et al., 2017), which is outside the scope of this work. In the unstructured biological model, biomass growth, product formation, maintenance, and decay processes (of biomass, products, and precursors) are described. The growth of biomass is written with the stoichiometric equation (Eq. (1.1)), where biomass ($CH_aO_bN_cP_d$), a product (P), and carbon dioxide (CO_2) are typically obtained from the conversion of the substrate (e.g., Glucose CH_2O), oxygen (O_2), a nitrogen (e.g., NH_3), and phosphate source (e.g., H_3PO_4) (Villadsen et al., 2011).

$$CH_2O + Y_{O/S} \cdot O_2 + Y_{N/S} \cdot NH_3 + Y_{Phos/S} \cdot H_3PO4 \rightarrow \\ Y_{X/S} \cdot CH_aO_bN_cP_d + Y_{P/S} \cdot P + Y_{C/S} \cdot CO_2 + (\cdots) + Y_{W/S} \cdot H_2O \quad (1.1)$$

Biomass is composed of many elements: C, H, O, N, P, S, and trace elements, such as Ca and Mg, which can be evaluated by elemental analysis. Only the dominant elements are usually considered for modelling purposes, and a pseudo steady state of this composition is assumed even though the elemental composition of the cell might slightly change during the process. The yield coefficients define the quantity of what is produced/consumed per quantity of consumed substrate, e.g., the biomass yield coefficient, $Y_{X/S}$, that describes how much biomass is produced per consumed substrate. The kinetic rates of the biological processes have been described in the literature using empirical relations, such as the Monod model (Monod, 1949) combined with functions accounting for inhibition and limiting effects of substrates, metabolites, products, and process conditions, like the pH (Eq. (1.2)).

$$\frac{dC_X}{dt} = \mu_{max} \cdot \frac{C_S}{C_S + K_S} \cdot \frac{K_P}{C_P + K_P} \cdot e^{-\left(\frac{(|pH_{opt} - pH|)^2}{\sigma_{pH}^2}\right)} \cdot C_X \quad (1.2)$$

where $\frac{dC_X}{dt}$ is the change of the biomass concentration (C_X) over time (t), μ_{max} is the maximum specific growth rate, C_S is the substrate concentration,

K_S the saturation parameter of the substrate, C_P the product concentration, K_P the product inhibition parameter, pH_{opt} the optimal pH for growth, and σ_{pH} the spread parameter of the Gaussian pH function. The latter approximates a bell-like curve relating the effect of pH on the maximum growth of cells.

Chemical processes are considered in bioprocess modelling in order to extend the application of models to systems, e.g., with varying pH. These models are applied to study for example the effect of the pH or precipitation on the process. Several chemical processes are modelled, including dissociation reactions of weak and strong acids and bases, ion pairing, and precipitation (Musvoto et al., 2000). Ion pairing and precipitation involve solid-liquid interactions, while a mixed acid/base system describes liquid-liquid interactions. In these processes, the analytical concentrations need to be adjusted by the activity coefficients (γ_i), owing to the changing interactions of ions in solution with each other and with the H_2O molecules at different ion concentrations. Various empirical equations are made use of for such activity corrections depending on the ionic strength (function of the concentration of all charged components) of the mixture (Loewenthal et al., 1989; Musvoto et al., 2000). As an example, Eq. (3) describes the dissociation reaction rate of an undissociated acid (HA) to its dissociated form (A^-) and the hydrogen ion (H^+):

$$r_{HA \rightleftharpoons A^- + H^+} = K'_{r,A} \cdot 10^{-\frac{pK_A}{\gamma_A^2}} \cdot C_{HA} - K'_{r,A} \cdot C_{A^-} \cdot C_{H^+} \tag{1.3}$$

where $K'_{r,A}$ is the apparent reverse rate constant for the acid dissociation, pK_A the acid dissociation constant for the specific acid considered, and C_i represents the concentration of the i-th component.

Physical processes deal with mass and heat transfer processes. Gas-liquid exchange processes are of special interest in cultivations. These include, among others, the mass transfer of the oxygen from the gas bubbles to the liquid broth (aeration) and the gas stripping of CO_2 from the liquid to the gas phase. Oxygen transfer is key for the success of most large-scale bioprocesses as most of them are aerobic – with the exception of, e.g., bioethanol and lactic acid production. Gas-liquid mass transfer models are usually based on the two-film theory (Whitman, 1962), and can be described as the product of the

volumetric mass transfer coefficient, k_La, and the driving force (concentration difference of the component in the liquid phase (CO_2) and the gas-liquid interface(C^*_{O2}) (Eq. (1.4)). There are numerous empirical correlations for k_La (Markopoulos et al., 2007), and several experimental methods have been developed to measure gas-liquid mass transfer (Villadsen et al., 2011).

$$r_{O_2(g)\rightleftharpoons O_2(l)} = k_L a_{O_2} \cdot (C^*_{O_2} - C_{O_2}) \tag{1.4}$$

Complex, non-linear models incorporate various biological, chemical, and physical processes. Detailed models include countless processes in order to better comprehend, for example, the anaerobic digestion process (Flores-Alsina et al., 2016), the pharmaceutical production of penicillin by *Penicillium chrysogenum* (Goldrick et al., 2015), different aeration and agitation conditions for enzyme production with *Aspergillus oryzae* (Albaek et al., 2011), antibiotic production with *Streptomyces coelicolor* (Sin et al., 2008), and *Pichia pastoris* and *Saccharomyces cerevisiae* fed-batch cultivations (Fernandes et al., 2012; Wechselberger et al., 2010). Many mechanistic models rely on a similar model structure based on first-principles chemical and physical process descriptions that have been added to the description of the biological mechanisms. This is because of the flexible structure of mechanistic models that allows them to be adapted easily to other bioprocesses." (**Paper A**)

Lactic Acid Bacteria Models

This study applied a mechanistic and unstructured model to model lactic acid bacteria cultivations, and this type of model is therefore reviewed. Various unstructured models have been applied to model lactic acid bacteria cultivations (Bouguettoucha et al., 2011). The growth kinetics and production of key metabolites (e.g., lactic acid for homofermentative lactic acid bacteria) are thus of primary interest.

Cell growth is generally described as:

$$\frac{dC_X}{dt} = \mu \cdot C_X \tag{1.5}$$

where μ represents the specific growth rate.

The biomass growth is affected by many factors, and typically carbon limitation and inhibition, product inhibition, temperature, and pH are considered in biokinetic models. μ is then modelled as the product of functions that account for these factors.

Carbon limitation is often described by the Monod model (Aghababaie et al., 2015; Zacharof and Lovitt, 2013; Altiok et al., 2006; Boonmee et al., 2003; Schepers et al., 2002a,b; Burgos-Rubio et al., 2000; Fu and Mathews, 1999):

$$\mu = \mu_{max} \cdot \frac{C_S}{C_S + K_S} \tag{1.6}$$

Carbon inhibition (Åkerberg et al., 1998; Altiok et al., 2006) could also be included with the Haldane equation as, e.g., Venkatesh et al. (1993) reported 60 g L^{-1} lactose as the inhibiting concentration (K_I) for *Lactobacillus bulgaricus*:

$$\mu = \mu_{max} \cdot \frac{C_S}{C_S + K_S + \frac{C_S^2}{K_I}} \tag{1.7}$$

Product inhibition refers to the inhibition by lactic acid that is produced in large quantities by lactic acid bacteria. Frequently non-competitive inhibition is assumed (Pinelli et al., 1997; Ohara et al., 1992; Cachon and Diviès, 1993):

$$f_P = \frac{K_P}{C_P + K_P} \tag{1.8}$$

A number of studies consider a critical lactic acid concentration (P_{inh}), above which growth is completely inhibited (Burgos-Rubio et al., 2000; Monteagudo et al., 1997; Youssef et al., 2005):

$$f_P = 1 - \frac{C_P}{P_{inh}} \tag{1.9}$$

Furthermore, various authors have modelled the effect of the undissociated (lactic acid: HLA) and dissociated (lactate: LA) forms separately (Schepers et al., 2002b; Aghababaie et al., 2015; Venkatesh et al., 1993; Amrane and Prigent, 1998). Aghababaie et al. (2015) applied, for instance, Eq. (1.10) and (1.11) to model *Streptococcus thermophilus* cultivations.

$$f_{HLA} = \frac{1}{1 + e^{K_P \cdot (C_{HLA} - K_{HLA})}} \tag{1.10}$$

$$f_{LA} = e^{-K_{LA} \cdot C_{LA}} \tag{1.11}$$

To model the effect of temperature, a modified Arrhenius law is generally applied (Esener et al., 1981; Aghababaie et al., 2015).

The effect of pH is usually modelled by a bell curve, whereas Schepers et al. (2002b) applied Eq.(1.12) with three parameters and Aghababaie et al. (2015) Eq. (1.13) with four parameters.

$$f_{pH} = e^{-\left(\frac{(|pH_{opt} - pH|)^n}{\sigma_{pH}^2}\right)} \tag{1.12}$$

$$f_{pH} = \frac{K_1 \cdot (pH_{opt} - pH)^2 + K_2}{(pH_{opt} - pH)^2 + K_3} \tag{1.13}$$

where pH_{opt} is the optimal pH, n the power parameter, σ the spread parameter, and $K_{1\text{-}3}$ coefficients.

Lactic acid production is mostly modelled by the Luedeking and Piret equation (Luedeking and Piret, 1959), as, for example, applied by Boonmee et al. (2003); Altiok et al. (2006); Burgos-Rubio et al. (2000); Åkerberg et al. (1998); Biazar et al. (2003); Rogers et al. (1978):

$$\frac{dC_P}{dt} = (\alpha \cdot \mu + \beta) \cdot C_X \tag{1.14}$$

where α represents the growth-associated production coefficient, and β the non-growth-associated coefficient.

Usually, the Luedeking and Piret equation is further modified, e.g., with a substrate-limitation term (Rogers et al., 1978):

$$\frac{dC_P}{dt} = \left(\alpha \cdot \mu + \beta \cdot \frac{C_S}{C_S + K_S}\right) \cdot C_X \tag{1.15}$$

A number of authors (Peng et al., 1997) have considered lactic acid production to be entirely growth dependent:

$$\frac{dC_P}{dt} = (\alpha \cdot \mu) \cdot C_X \tag{1.16}$$

Altiok et al. (2006) summarized the estimated kinetic parameters (μ_{max}, K_S, α, $Y_{X/S}$ etc.) of many lactic acid bacteria studies. They depend, of course,

on the strain and cultivation conditions but they serve as a robust indication, e.g., for future parameter estimations.

Applications of Process Models

"Mechanistic models can be applied to a range of tasks from the stage of process development to implementation during industrial-scale production (Mears et al., 2017b) – they are applied, for instance, to process development, optimization, monitoring, and control. The applications contribute to economic and sustainable production and comply also with PAT guidelines (Gernaey et al., 2012).

Mechanistic models are applied during the process development phase to design a new process. Design means typically that a new process is formulated for an existing plant as it is rather rare that a new production plant is constructed for a novel process. This process involves the definition of optimal process operating conditions, such as stirrer speed, substrate-addition rate (for fed-batch cultivations), and aeration rate under the given constraints of the available equipment as demonstrated for enzyme production with *Aspergillus oryzae* (Albaek et al., 2011). Initial biotechnological strain and process development is carried out at the laboratory scale, which is further refined and validated under pilot-scale conditions. The final process is then scaled up to the production scale because experiments at the large scale are rather expensive. Mechanistic models are especially attractive at this stage – it is often the case that very little data is available, and they are capable of testing new conditions or equipment for which there is little or no data available. The models can help assess and understand the effects of large-scale process conditions as process scale-up is very challenging (Neubauer et al., 2013; Stocks, 2013; Wang et al., 2014). Traditionally, bioprocess scale-up is based on physical parameters, such as volumetric power input, the volumetric oxygen mass transfer coefficient, k_La, or stirrer tip speed. However, biological properties and the effect of the heterogeneous cultivation environment (see subsection: Gradients at a Large Scale) are often not taken into account. Another challenge is varying process conditions, such as humidity, aeration rate, and impeller speed or type, when different bioreactors are used at different production sites all over the world, but the

same product quality is of course demanded.

Optimizing an established process is key for the long-lasting success of a manufacturer in order to withstand economic pressures in a competitive market. Mechanistic models have been successfully applied to optimize cultivations for many processes. As an example, Jiménez-Hornero et al. (2009) proposed a model-based optimization for acetic acid cultivation. Models are also applied to optimize not the main cultivation itself, but the pre-cultures instead. Frahm (2013) optimized the seed train for biopharmaceutical production. The seed is required to inoculate the production bioreactor with the right amount of cells in order to allow for stable production. Toumi et al. (2010) employed multi-unit process simulation in order to optimize large-scale monoclonal antibody production.

Real-time determination of CQAs is of utmost interest in the biotechnological industry. However, just a few probes are available that measure the required attributes in the bioreactor (in-line). Biomass can, for example, be measured indirectly in-line by capacitance or turbidity measurements. However, changing conditions of the complex biological process matrix influence these measurement techniques, and make a reliable prediction challenging. For the same reason, probes measuring metabolite concentrations (e.g., product concentrations) are challenging, as well. Further reasons for the current lack of in-line probes are that they must be robust, give stable signals, and withstand harsh conditions during sterilization and cleaning procedures. In addition, they must fulfil regulatory requirements according to good manufacturing practices (GMPs).

Software sensors are an established alternative to hardware probes for the monitoring of bioprocesses and countless implementations in the bioprocess industry can be found (Biechele et al., 2015; Luttmann et al., 2012; Mandenius and Gustavsson, 2014; Pais et al., 2014; Posch et al., 2013; Sagmeister et al., 2013; Zhao et al., 2015). Software sensors are often also called soft sensors or state estimators. Soft sensors utilize the available on-line measurements, such as the exhaust gas analysis results (e.g., O_2, CO_2, and other volatile compounds), pH, temperature, pressures, flow rates, and stirrer speed. These measurements are then used as input for the model that

predicts the unknown process states including the CQA. In addition to the aforementioned measurements, spectroscopic methods are sometimes applied in combination with data-driven models (Kadlec et al., 2009). Spectroscopic methods include UV-Vis (ultraviolet–visible) (Ödman et al., 2009; Zavatti et al., 2016), near-infrared (Alves-Rausch et al., 2014), and mid-infrared fluorescence spectroscopy (Fayolle et al., 1997), and Raman spectroscopy (Golabgir and Herwig, 2016) among others. Automated microscopy for acquiring cell-specific information, such as cell morphology, are intensively studied for monitoring (Marquard et al., 2017) and control strategies (Bluma et al., 2010), e.g., for the bioethanol production process (Belini et al., 2013).

Models are also applied for control strategies within the bioprocess industry. Bioprocesses are often fed-batch processes, where a feed solution is added continuously to the cultivation broth. Models are applied, for instance, to define the feed flow rate in order to control substrate concentration in the broth. As there are no sufficient hardware probes to assess the substrate concentration in real-time, models are used to predict the substrate concentration to enable better control strategies (Craven et al., 2014; Johnsson, 2015; Mears et al., 2017a).

1.3 Uncertainties in Model Predictions

Owing to the complexity of bioprocesses and mechanistic models, it is necessary to consider uncertainties when a model is applied. There are three sources of uncertainties: i) stochastic uncertainty; ii) input uncertainty; and iii) structural uncertainty (McKay et al., 1999):

> i) Stochastic uncertainty covers the stochastic variabilities that are observed in real processes, for example random failure of equipment that leads to a disturbance in the process. The unforeseeable shutdown of air sparging, as an example, would result in a shortage of oxygen.
>
> ii) The input uncertainties refer to the lack of perfect knowledge surrounding the model parameters and model inputs. The model parameters rely on available measured data, which are naturally subject to random measurement errors. A parameter value is characterized by a

probability distribution around its nominal value instead of considering a single value. In addition to the uncertain model parameters, model inputs, such as initial process conditions and on-line measurements, are uncertain. The initial substrate concentration, for example, varies from batch to batch owing to variations in the medium preparation process, hence a fixed value cannot be expected.

iii) The structural uncertainties are related to the mathematical description of the model. We typically have insufficient knowledge of the bioprocess that would allow a model description, including all relevant details of the process. Our applied models are therefore an approximation of the process and based on assumptions. In order to analyse the effect of the uncertainties, an uncertainty analysis is performed. This is an important element of good modelling practice to ensure a reliable mechanistic model and robust PAT application (Sin et al., 2009a). During the uncertainty analysis, the uncertainties are propagated to the model outputs and their effect is evaluated. There are certain methods for uncertainty analysis available, including Monte Carlo procedure, differential analysis, response surface methodology, the Fourier amplitude sensitivity test, and Sobol variance decomposition (Helton and Davis, 2003; Saltelli et al., 2008). In this contribution, we focus on the Monte Carlo procedure as this method is widely used and reliable (Helton and Davis, 2003).

Probabilistic Model Predictions

The uncertainty analysis provides probabilistic information about the performance of the bioprocess. It shows, for example, what the probability is that the required product quantity or quality will be achieved by considering the aforementioned input uncertainties. Frequently, the Monte Carlo procedure is used to propagate the input uncertainties to the output predictions (Sin et al., 2009a). The Monte Carlo procedure involves three steps: 1) defining the input uncertainties; 2) sampling within the input uncertainty; and 3) Monte Carlo simulation to obtain the model output uncertainty.

In step 1, the model input uncertainties are defined, which is typically

conducted based on expert opinions. This involves both asking process experts from industry and academia as well as consulting relevant literature. The aim is to obtain the uncertainty for model parameters, initial process conditions, and measurements, e.g., what is the lower and upper bound of these variables. A probability distribution is then assigned to the variables. Often, a uniform distribution, also known as uninformed prior, is assumed when there is a lack of observations or data to support the derivation of a specific distribution function, such as normal distribution. The latter is usually assumed to describe the distribution function for measurement errors. For the model parameters, the maximum likelihood estimation theory can be employed to derive a multivariate normal distribution and the covariance matrix. Such information should then be used for appropriate definition of input model parameter uncertainties (Sin et al., 2010).

In step 2, random combinations of the model inputs are sampled considering the previously defined uncertainties (Fig. 1.3, Input samples). As an example, the parameters, A and B, representing model input parameters are considered to be uncertain, and N=1000 input samples are sampled. There exist several sampling methods, including random sampling (Meng, 2013), low-discrepancy sequence, such as Halton (Halton, 1964) and Sobol (Sobol', 1967) sequences, and stratified sampling, including Latin Hypercube Sampling (LHS) (McKay et al., 1979). The sampling method is used to sample N independent inputs from the probability domain [0 1], where N is the number of input samples. Most of the time, parameters and inputs are correlated (see the covariance matrix estimation from the maximum likelihood theory). To preserve the correlation, dependent samples need to be generated, e.g., by applying the multivariate probability distribution copula (Nelsen, 2006) or the Iman-Conover method (Iman and Conover, 1982). The Iman-Conover method employs a rank-based correlation control method to induce the desired correlation between the parameter samples from an independent input space. For those parameters, initial conditions, and measurements that lack correlation information, an identity correlation matrix is used (meaning no correlation is assumed). Finally, the input

samples are inverted from the probability domain [0 1] to real values with an inverse cumulative distribution function corresponding to the distribution function for each input.

In step 3, the input uncertainties are propagated using the Monte Carlo procedure to estimate output uncertainty (Fig. 1.3, Monte Carlo simulation). The model is thereby simulated N times with each of the defined input sample sets.

The results are N predictions of the model outputs (Fig. 1.3, Model outputs). For each model output, the span of the model output prediction indicates the extent of its uncertainty during the cultivation. Inferential statistical analysis such as mean, and 90^{th} and 10^{th} percentiles are applied to assess the results. It is important to realize that the results of the uncertainties can only be interpreted in the analysis boundaries/frame (Sin et al., 2009b). This is because the definitions and assumptions that are made in the study, e.g., the selected model, the identification and characterization of uncertainties, and the selected methodology.

Many studies have investigated the effect of uncertainties on the model outputs for different biotechnological processes, including a hydrothermal pre-treatment process of lignocellulosic biomass (Prunescu et al., 2015), a milk-drying process (Ferrari et al., 2016), and antibiotic production with *Streptomyces coelicolor* (Sin et al., 2008).

Risk-based Decision Making

Bioprocesses at the production scale undergo a risk assessment to ensure the quality of manufacturing. Risk assessment seeks to increase the safety of the process along with the quality of the product. Most commonly, risk-management tools recommended by ICH Q9 (2005) are used, such as the Failure Mode and Effects Analysis (FMEA). In this worksheet-based method, the risks related to e.g., material properties and process conditions, are quantified first. This characterization is fundamental to understanding the impact CPPs on CQAs. Then, the most important critical process parameters are prioritized and further experiments could be carried out if needed. Finally,

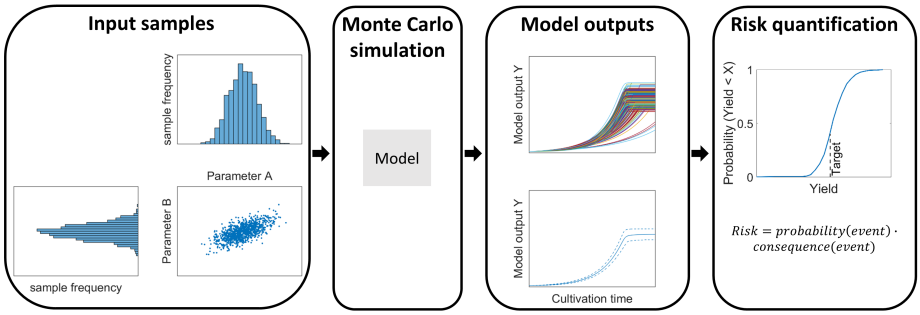


Figure 1.3: Risk quantification using the Monte Carlo procedure with four steps. 1. Input uncertainty space and sampling; 2. Monte Carlo simulation; 3. Model outputs; 4. Risk quantification. Reprinted from **Paper A**.

the design space for production is defined, i.e., the acceptable variability, e.g., in material properties and process conditions is defined. The process operation within the design space minimizes then the risks of obtaining faulty batches. Undesired events are minimized through controlling the critical process parameters within the design space. Furthermore, preventative repair and maintenance actions are performed, such as the inspection and calibration of equipment. Published examples of this Quality by Design (QbD) approach include the mixing unit operation (Adam et al., 2011) and spray drying process (Baldinger et al., 2011).

Mechanistic simulations are however rarely used for risk assessment in the biotechnological industry (Rantanen and Khinast, 2015) even though PAT applications that consider various uncertainties provide probabilistic model predictions. They can therefore support risk-based decisions in real time (Stocker et al., 2014). A probabilistic bioprocess model predicts the probability distribution of CQAs and enables therefore the operators to react on a risk-based basis. For example, the cumulative distribution function that indicates the probability of not achieving the target (e.g., target yield) can be derived from the Monte Carlo simulation outputs (Fig. 1.3, Step 4: Risk quantification). The risk is then quantified as the product of the probability of the undesired event times the consequence of the undesired event (Cameron and Raman, 2005). The risk could be, e.g., a loss of 0.5 kg product per cultivation or in economic perspectives e.g., $ 0.2 million.

This then contributes to risk-based monitoring and controlling, and thereby improves both the performance and safety of the production process (Li et al., 2008).

Both the Bayesian approach (Mockus et al., 1997; Peterson and Lief, 2010) and Monte Carlo approach (García-Muñoz et al., 2015) have been investigated and recommended for probabilistic PAT applications (Sin et al., 2009a; Tabora and Domagalski, 2017). Depending on the purpose of the application, stochastic uncertainties might be incorporated into model application. The probability of a specific event occurring, e.g., pump failure, could then be integrated (Barua et al., 2016), hence engineers could assess the control algorithm robustness and evaluate, for example, the response of the controller in the case of a pump failure (Sin et al., 2009a). Konakovsky et al. (2017) optimized the glucose feeding strategy for Chinese hamster ovary (CHO) cells in fed-batch culture, and considered various uncertainties in a Monte Carlo simulation assuming a worst-case scenario *in silico*. However, there is a considerable lack of published risk-based monitoring and control implementations for large-scale cultivations, especially using mechanistic models, despite their potential. One major challenge might be the time that is needed to develop and implement the model-based control system. A second challenge might be regulatory validation requirements (Djuris and Djuric, 2017)." (**Paper A**)

Chapter 2

Model Development and Validation

As the basis of this work, this chapter describes the development of the mechanistic model that was applied to simulate *Streptococcus thermophilus* batch cultivations. The model was fitted against experimental lab-scale data and validated with independent batch and continuous experiments. The target product of the studied lactic acid bacteria cultivation is the biomass that is produced as starter cultures for the dairy industry, e.g., to produce cheese or yogurt.

This chapter is based upon the following article:

Paper B: Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. Biochemical Engineering Journal, 135, 49–60.

2.1 Model Development

A mechanistic model was developed with the aim of describing *Streptococcus thermophilus* cultivations. It was an unstructured unsegregated dynamic model that comprised a biokinetic model and chemical model. The biokinetic model described the evolution of the biological state variables, such as biomass, the substrate lactose, and the side product lactic acid of the *S. ther-*

mophilus cultivation. The chemical model, also called mixed weak acid/base model, covered the dissociation reactions of the charged components in the cultivation broth in order to predict the pH of the cultivation as the pH was of major interest in this work to account for pH gradients at large scales. The model was implemented and solved in MATLAB® R2017a (The MathWorks®, Natick, MA) using the solver ode15s.

Biokinetic Model

The dynamic biokinetic model was based on overall process stoichiometry (Villadsen et al., 2011) (Eq. (2.1)). The carbon source lactose is consumed together with ammonia and phosphoric acid, and converted to biomass, lactic acid, and galactose. As an elemental analysis of the *S. thermophilus* strain was not available, a biomass composition of $CH_{1.95}\ O_{0.63}\ N_{0.22}\ P_{0.02}$ was assumed in the present study (Oliveira et al., 2005), which was originally determined for *Lactococcus lactis*.

$$q_S \cdot CH_2O + q_{NH} \cdot NH_3 + q_{Ph} \cdot H_3PO_4 \rightarrow q_X \cdot CH_{1.95}O_{0.63}N_{0.22}P_{0.02} + q_P \cdot CH_2O + q_{Gal} \cdot CH_2O \quad (2.1)$$

The biomass growth rate was modelled as a product of functions. These functions had values between 0 and 1, and accounted for the lag-time (f_{lag}), lactose inhibition and limitation (f_S) (Åkerberg et al., 1998), lactate inhibition(f_P) (Aghababaie et al., 2015), and the pH in the cultivation broth (f_{pH}) (Schepers et al., 2002a) (Eqs. (2.2 - 2.3)). According to the studies of Schepers et al. (2002a) and Amrane and Prigent (1998), the dissociated form of lactic acid was growth inhibiting under the investigated pH conditions.

$$\frac{dC_X}{dt} = \mu_{max} \cdot f_{lag} \cdot f_S \cdot f_P \cdot f_{pH} \cdot C_X \quad (2.2)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot (1 - e^{-t/t_{lag}}) \cdot \frac{C_S}{C_S + K_S + \frac{C_S^2}{K_I}} \cdot \frac{1}{1 + e^{K_{P,La} \cdot \left(C_{La} - K_{La} \cdot \frac{1}{1+e^{K_{P,pH1} \cdot (pH - K_{P,pH2})}} \right)}} \cdot e^{-\left(\frac{(pH_{opt} - pH)^2}{\sigma_{pH}^2} \right)} \cdot C_X \quad (2.3)$$

The lactic acid synthesis was assumed to be growth dependent (Peng et al., 1997):

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} \quad (2.4)$$

The lactose consumption was the sum of biomass growth and lactic acid synthesis rate. The studied strain metabolized only glucose and secretes galactose with the yield (Y_{Gal}: galactose consumed per lactose) under the present cultivation conditions:

$$\frac{dC_S}{dt} = -(1 + Y_{Gal}) \cdot \left(\frac{dC_X}{dt} + \frac{dC_P}{dt} \right) \quad (2.5)$$

The stoichiometric matrix of the biokinetic model component is seen in Table 2.1.

Table 2.1: Stoichiometric matrix of the biokinetic model.

Component name	Lactose	Galactose	Biomass	Lactate	Ammonia	Phosphoric acid
Chemical composition	$C_{12}H_{22}O_{11}$	$C_6H_{12}O_6$	$CH_{1.95}O_{0.63}N_{0.22}P_{0.02}$	$C_3H_5O_3^-$	NH_3	H_3PO_4
Process ↓	C-mol L^{-1}	C-mol L^{-1}	C-mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}
Biomass growth	-(1+Y_{gal})	Y_{gal}	1		-0.22	-0.02
Lactic acid synthesis	-(1+Y_{gal})	Y_{gal}		$\frac{1}{3}$		

Note. Reprinted from **Paper B**.

Table 2.2: Stoichiometric matrix of the mixed weak acid/base model.

Component name	Lactic acid	Lactate	Ammonia	Ammonium	Phosphoric acid	Dihydrogen phosphate	Hydrogen phosphate	Carbon dioxide (dissolved)	Bicarbonate	Hydrogen ion	Hydroxyl ion	unknown base dissociated	unknown base undissociated
Chemical composition	$C_3H_6O_3$	$C_3H_5O_3^-$	NH_3	NH_4^+	H_3PO_4	$H_2PO_4^-$	HPO_4^{2-}	$H_2CO_3^*$	HCO_3^-	H^+	OH^-	Z^+	Z
Process ↓	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}	mol L^{-1}
Ammonia dissociation			1	-1						1			
Phosphate diss. 1					-1	1				1			
Phosphate diss. 2						-1	1			1			
Carbonate diss. 1								-1	1	1			
Lactate diss.	-1	1								1			
Water diss.										1	1		
Dissociation of Z										1		-1	1

Note. Reprinted from **Paper B**.

Mixed Weak Acid/Base Model

The goal of the mixed weak acid/base model was to predict pH during the cultivation. This model component comprised the dissociation reactions of the charged compounds in the cultivation broth (Musvoto et al., 2000) (Table 2.3). The dissociation reactions of lactic acid, ammonia, phosphoric acid, carbonic acid, water, and an unspecified compound Z were considered. Z was introduced to account for the unknown compounds in the cultivation broth, such as amino acids, as a complex medium was used and not all components were identified and quantified.

The kinetic parameters of the mixed weak acid/base model can be found in **Paper B**. In order to account for the non-ideal interactions of the charged components, the activity coefficients (f_i) were calculated by a modified form of the Debye-Hückel theory from Davies (1962):

$$log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3 \cdot I \right) \tag{2.6}$$

with the ionic strength (I):

$$I = \frac{1}{2} \sum_i (z_i^2 \cdot C_i) \tag{2.7}$$

where i are all charged components.

Table 2.3: Kinetics for the mixed weak acid/base model.

reaction	reaction rate vector
$NH_4^+ \rightleftharpoons NH_3 + H^+$	$K'_{r,NH} \cdot K'_{NH} \cdot C_{NH_4^+} - K'_{r,NH} \cdot C_{NH_3} \cdot C_{H^+}$
$H_3PO_4 \rightleftharpoons H_2PO_4^- + H^+$	$K'_{r,P1} \cdot K'_{P1} \cdot C_{H_3PO_4} - K'_{r,P1} \cdot C_{H_2PO_4^-} \cdot C_{H^+}$
$H_2PO_4^- \rightleftharpoons HPO_4^{2-} + H^+$	$K'_{r,P2} \cdot K'_{P2} \cdot C_{H_2PO_4^-} - K'_{r,P2} \cdot C_{HPO_4^{2-}} \cdot C_{H^+}$
$H_2CO_3^* \rightleftharpoons HCO_3^- + H^+$	$K'_{r,C1} \cdot K'_{C1} \cdot C_{H_2CO_3^*} - K'_{r,C1} \cdot C_{HCO_3^-} \cdot C_{H^+}$
$C_3H_6O_3 \rightleftharpoons C_3H_5O_3^- + H^+$	$K'_{r,LA} \cdot K'_{LA} \cdot C_{C_3H_6O_3} - K'_{r,LA} \cdot C_{C_3H_5O_3^-} \cdot C_{H^+}$
$H_2O \rightleftharpoons OH^- + H^+$	$K'_{r,W} \cdot K'_W \cdot C_{H_2O} - K'_{r,W} \cdot C_{OH^-} \cdot C_{H^+}$
$ZH^+ \rightleftharpoons Z + H^+$	$K'_{r,Z} \cdot K'_Z \cdot C_{ZH^+} - K'_{r,Z} \cdot C_Z \cdot C_{H^+}$

Note. Adapted from **Paper B**.

Nevertheless, the validated range I $\leq$ 0.5 mol L^{-1} of the applied Davis equation to calculate the activity coefficients (Eq. (2.6)) has to be noted, and could be improved in future studies in particular as the ionic strength was occasionally higher than 1 mol L^{-1} in the cultivations.

The stoichiometric matrix of the mixed weak acid/base model part is shown in Table 2.2.

A P-controller with a controller gain (K_P) of 5 $mol\ L^{-1} \cdot Volume\ [L]$ was implemented in the simulations to maintain the pH at the set point value by adding ammonia solution (Eq. (2.8)).

$$NH_4OH_{add} = K_P \cdot (pH_{set} - pH) \tag{2.8}$$

2.2 Cultivation Conditions

Streptococcus thermophilus batch cultivations were performed in well-mixed 2-L stirred-tank bioreactors (Biostat® B, Sartorius AG, Germany) for the model development and validation (this chapter) and the application of the soft sensor (Chapter 3). The 2-L bioreactors were operated at 300 rpm. The cultivation temperature was 40 °C, and the headspace of the bioreactor was gassed with nitrogen. The pH was controlled by adding 24 % ammonia solution to the top of the liquid phase. The cultivation medium contained 20 or 65 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$. The pH and ammonia addition (balance value readout) were measured on-line. Sugars and organic acids were quantified off-line from filtered samples (filter pore size: 0.2 μm) in a high-performance liquid chromatography (HPLC) system as described in detail in **Paper B**.

2.3 Parameter Estimation

A parameter estimation was performed to estimate the kinetic parameters of the biokinetic model. The model was fitted to the experimental lactose, biomass, and lactic acid concentrations of five cultivations. These batch

Table 2.4: Initial parameter values of the biokinetic model.

kinetic parameters	initial value	source of initial value	lower bound	upper bound
μ_{max}	1.18 h^{-1}	Aghababaie et al. (2015)	1	5
K_S	0.79 g L^{-1}	Åkerberg et al. (1998)	0.5	5
K_I	164 g L^{-1}	Åkerberg et al. (1998)	50	200
$K_{P,La}$	0.2 L g^{-1}	Schepers et al. (2002a)	0.1	1
K_{La}	25 g L^{-1}	Schepers et al. (2002a)	10	30
pH_{opt}	6	expert knowledge	5.7	6.5
σ_{pH}	1.54	Schepers et al. (2002a)	0.1	3
α	5 g g^{-1}	expert knowledge	0.1	30
$K_{P,pH1}$	20	expert knowledge	1	50
$K_{P,pH2}$	7	expert knowledge	6.8	7.2
$pH_{opt,lag}$	5.7	expert knowledge	5.5	6
σ_{lag}	0.4	expert knowledge	0.3	0.5
Y_{gal}	0.6 g g^{-1}	expert knowledge	0	1

Note. Reprinted from **Paper B**.

cultivations were conducted with an initial lactose concentration of 65 g L^{-1} under different constant pH conditions: 1x pH 5.5, 2x pH 6.0, 1x pH 6.5, and 1x pH 7.0. The pH was held constant at the set point in the simulation for the parameter estimation, and the mixed weak acid/base model was not considered in order to obtain biokinetic parameter estimates that were independent of the mixed weak/acid base system. Initial parameter values were taken from the literature (Åkerberg et al., 1998; Aghababaie et al., 2015; Schepers et al., 2002a) (Table 2.4).

The parameter estimation was conducted in MATLAB with the nonlinear least-squares solver lsqnonlin, and this was followed by the maximum likelihood estimation method from Seber and Wild (1989) as described in detail in **Paper B** and **Paper C**. As part of the good modelling practice a comprehensive assessment of the model parameters was conducted (Sin et al., 2009a). This included an identifiability, sensitivity, and uncertainty analysis in order to garner knowledge of the limitations of the model and to find an identifiable parameter subset for regression.

The identifiability analysis revealed that nine parameters were identifiable.

Table 2.5: Estimated model parameters including the relative errors and correlation matrix.

kinetic parameters	estimated parameter value	relative error [%]	correlation matrix								
			μ_{max}	$K_{P,La}$	K_{La}	pH_{opt}	σ_{pH}	α	$pH_{opt,lag}$	σ_{lag}	Y_{gal}
μ_{max}	2.06	1	1	-0.74	-0.83	-0.17	0.5	-0.08	-0.52	0.53	0
$K_{P,La}$	0.24	13		1	0.77	-0.2	-0.58	-0.05	0.39	-0.28	0.04
K_{La}	19.80	0			1	-0.07	-0.54	0.31	0.44	-0.35	-0.28
pH_{opt}	6.39	1				1	-0.52	-0.13	0.76	-0.86	0.02
σ_{pH}	1.42	3					1	-0.08	-0.93	0.85	0.06
α	5.19	0						1	-0.1	0.12	-0.4
$pH_{opt,lag}$	5.70	1							1	-0.97	-0.03
σ_{lag}	0.3	9								1	0.03
Y_{gal}	0.69	5									1

Note. Reprinted from **Paper B**.

The estimated parameter values were in the anticipated order of magnitude known from the literature (Table 2.5). Only μ_{max} = 2.06 h^{-1} was higher than the actual biological value because it had to compensate for the functions in the growth rate expression. On average, each function has a value of around 0.9 during the cultivation, resulting in a biological μ_{max} = 1.35 h^{-1} ($0.9^4 \cdot 2.06$). Of course, the estimated parameters are subject to model structure, cultivation conditions, and nominal parameter values among others (Sin et al., 2010).

The relative error (RE), which is the ratio between the standard deviation of the estimated parameter and the mean value, was used to quantify the uncertainty of the parameter estimates (Eq. (2.9)). The relative error of all parameters was lower than 10 % (except $RE_{K_{P,La}}$ = 13 %), and hence acceptable (Table 2.5).

$$RE_i = \frac{\sigma_{\widehat{\theta}_i}}{\widehat{\theta}_i} \tag{2.9}$$

The evaluation of the correlations between the parameters revealed that certain parameters were uniquely identifiable (correlation coefficient < 0.5) (Table 2.5). However, there was a linear dependency between some other parameters. The estimated value of one parameter is therefore conditional upon the value of another parameter, and the parameter set must be

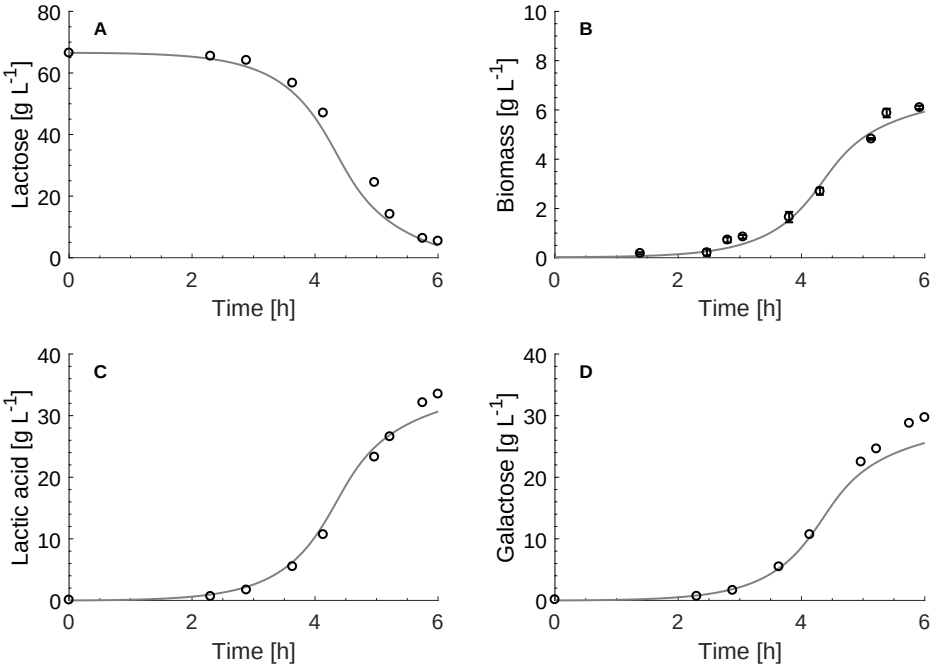


Figure 2.1: Model predictions for the *S. thermophilus* lab-scale batch cultivation that was employed for parameter estimation. Lactose (A), biomass with the standard deviation (B), lactic acid (C), and galactose (D) concentrations. The cultivation was performed in a 2-L stirred-tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. The model prediction (solid lines) for the measurements (circles). Reprinted from **Paper B**.

considered as a whole.

The model showed an acceptable fit of the cultivation data. Fig. 2.1 shows the model fit at pH = 6, while the model predictions under the other cultivation conditions can be found in **Paper B**. After 5-6 hours, 6 g L^{-1} biomass were yielded, and the stationary growth phase was reached (Fig. 2.1 B). Lactose was almost entirely consumed, while 5 g L^{-1} were left over (Fig. 2.1 A). A total of 33 g L^{-1} total lactic acid (lactic acid and lactate) were produced (Fig. 2.1 C). A total of 30 g L^{-1} galactose was measured in the cultivation broth at the end of the cultivation (Fig. 2.1 D). The goodness of fit for the model predictions was assessed with the root mean sum of squared error (RMSSE) (Eq. (2.10)), where the main interest was in the accuracy of the biomass concentration because this was the product of interest.

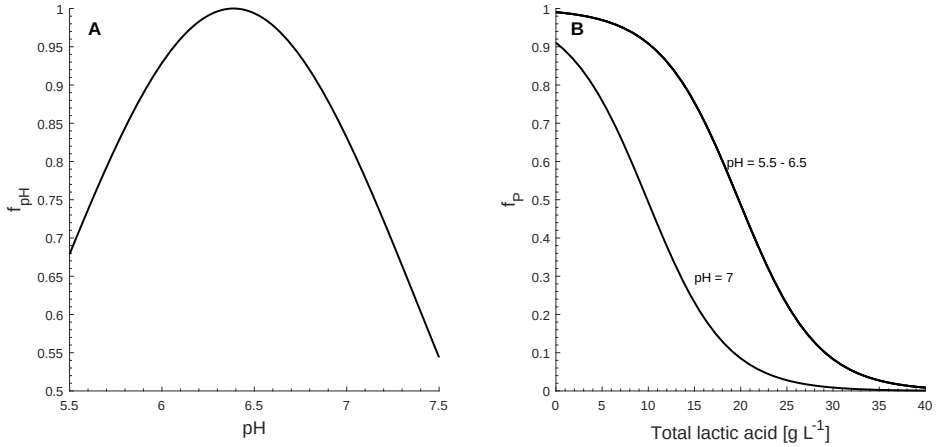


Figure 2.2: Growth affecting functions of pH and lactate inhibition. pH function (f_{pH}) vs. pH (A) and lactate inhibition function (f_P) vs. the total lactic acid concentration (B). Reprinted from **Paper B**.

$$RMSSE = \sqrt{\frac{1}{n} \cdot \sum_{i}^{n} (y_{meas,i} - \widehat{y}_i)^2} \tag{2.10}$$

The RMSSE for biomass was smaller than 0.6 g L^{-1} for the cultivations at pH 5.5, 6.0, and 6.5, corresponding to a discrepancy of less than 10 %, and giving evidence of a strong fit. The model fit at pH = 7.0 had an error of 30 %. The galactose concentration was slightly underestimated for all cultivations. This may be explained by the inconsistent carbon balance of the experimental cultivation data. There, more carbon was produced in the form of biomass, lactic acid, and galactose than carbon, which was derived from lactose, was consumed. In the experiments, yeast extract was supplemented but not considered in the model. It is likely that components of the yeast extract, such as amino acids, were taken up by the cells, and might therefore have led to inconsistency in the carbon balance.

Next, the functions, f_{pH} and f_P, were assessed. The bell-shape pH function f_{pH} had maximum at pH = 6.4 (Fig. 2.2 A). At pH = 5.5 and 7, growth was reduced by 25 %. Previous studies observed a similar influence of the pH on the growth of lactic acid bacteria even though slightly different pH optimums have been determined owing to different strains being studied (Aghababaie et al., 2015; Zacharof and Lovitt, 2013; Ohara et al., 1992).

The inhibition of growth by lactate (f_P) was pH dependent (Fig. 2.2 B). A total of 20 g L^{-1} total lactic acid inhibited growth by 50 % in a pH range from 5.5 to 6.5. At pH = 7, only 10 g L^{-1} total lactic acid inhibited growth by 50 % already. Nandasana and Kumar (2008) already noticed pH-dependent lactate (the major component under the given conditions) inhibition of the biomass growth for lactic acid-producing bacterium, *Enterococcus faecalis*. This emphasizes the requirement of the pH-dependent lactate inhibition function (Eq. (2.3)).

2.4 Model Validation

Batch Cultivation

Following parameter estimation, the model was validated with independent cultivation data sets. The cultivations were performed with an initial lactose concentration of 20 g L^{-1} at pH = 6.0 (Figs. 2.3 and 2.4). Here, the mixed weak acid/base model was also applied to predict pH. The model predicted the off-line measured lactose, biomass, total lactic acid, and galactose concentrations (Figs. 2.3 and 2.4 A-D). In addition, the pH and quantity of added ammonia were predicted (Figs. 2.3 and 2.4 E-F). An acceptable prediction accuracy was achieved for all state variables. The RMSSE for biomass was 0.2 g L^{-1} for both cultivations, which corresponds to an error of 10 % with respect to the final biomass concentration. The pH could be accurately predicted with a discrepancy of less than 0.1 pH units in the one case (Fig. 2.3 E) and less than 0.3 pH units in the other case (Fig. 2.4 E). When the cultivation started, the pH dropped from 6.1 to 6.0. Then, the pH was controlled by adding an ammonia solution. The initial drop of the pH was predicted to be faster than actually measured. This could be attributed to a different buffer capacity of the experimental medium in comparison to the modelled buffer capacity in the mixed weak acid/base model. In the second validation cultivation, a drift of the pH sensor might have caused the greater pH discrepancy. The obtained pH prediction accuracy was deemed sufficiently accurate for the purposes of this study seeing that measurement errors in the range of 0.1 pH unit were expected.

The model prediction of the pH was conditional on the predictions of

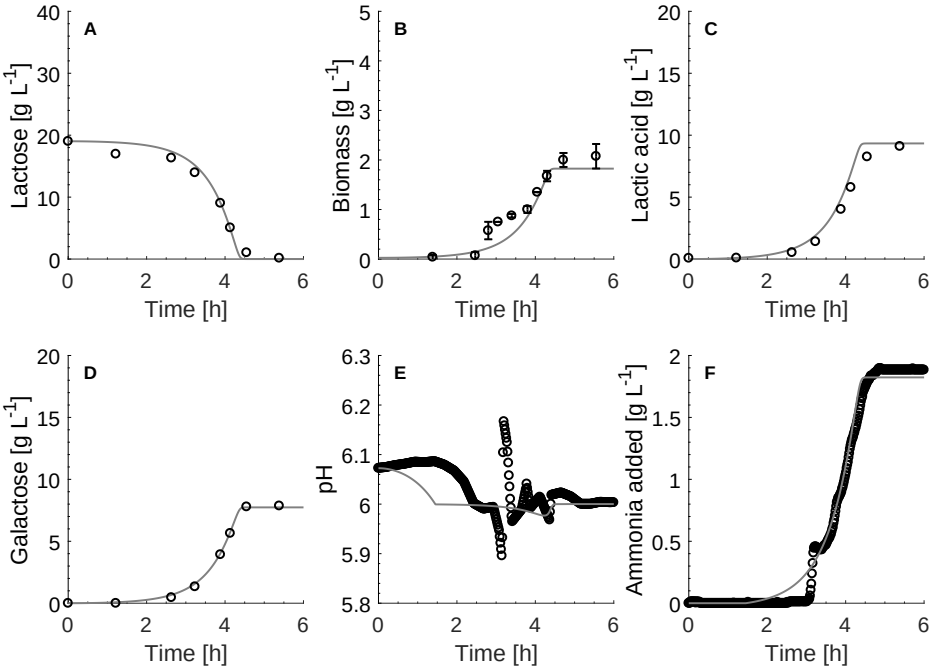


Figure 2.3: Model predictions of the validation cultivation 1. Lactose (A), biomass with standard deviation (B), lactic acid (C), galactose (D) concentrations, pH (E), and the quantity of added ammonia (F). Model prediction (solid line) and the measurements (circles). The *S. thermophilus* cultivation was performed in a 2-L stirred-tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. Reprinted from **Paper B**.

the lactic acid concentration and the added ammonia quantity, and hence the ammonia concentration in the broth. The validity of the mixed weak acid/base model could therefore not be demonstrated by a correct prediction of the pH as the pH is controlled in the model. The accurate prediction was rather demonstrated by the quantity of added ammonia (Figs. 2.3 and 2.4 F). As the model output matched the balance readout measurements of the base addition, the pH model was deemed to be accurate.

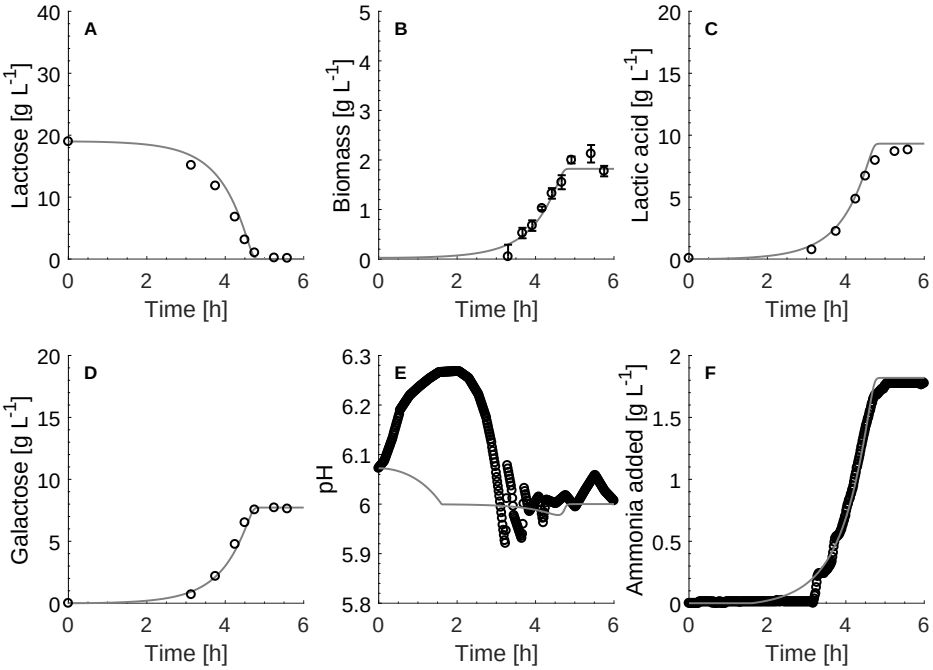


Figure 2.4: Model predictions of the validation cultivation 2. Lactose (A), biomass with standard deviation (B), lactic acid (C), galactose (D) concentrations, pH (E), and the added ammonia quantity (F). Model prediction (solid line) and the measurements (circles). The *S. thermophilus* cultivation was performed in a 2-L stirred-tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. Reprinted from **Paper B**.

Continuous Cultivation

The biokinetic model was also validated with a continuous accelerostat cultivation in addition to the previously discussed batch cultivations. The cultivation was performed in a 0.3 L bioreactor with a lactose concentration of 20 g L^{-1} both as the initial concentration in the batch phase and in the feed solution. The cultivation began with a batch phase that was followed by a chemostat at a dilution rate of first 0.3 h^{-1} for ca. 50 h, and then 0.1 h^{-1} for ca. 50 h (Fig. 2.5). Subsequently, the accelerostat was performed with an acceleration rate of the feed of 0.005 h^{-2}. Later on, the acceleration rate was increased stepwise to 0.008 h^{-2}. This experiment was performed in cooperation with Klaus Pellicer Alborch (Technische Universität Berlin, Germany).

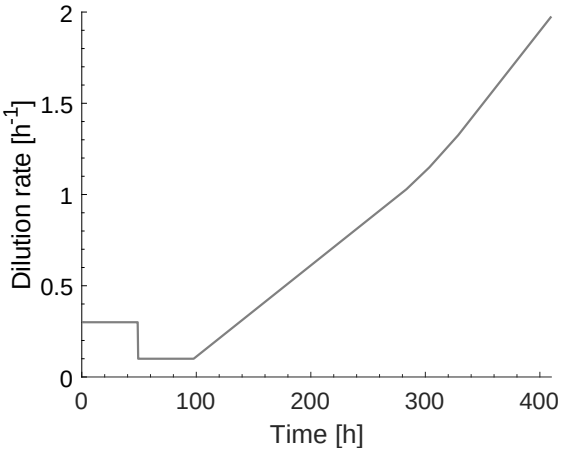


Figure 2.5: Dilution rate of the continuous cultivation (accelerostat). The cultivation started with a batch phase (not shown), followed by a chemostat and an accelerostat. The acceleration rate was 0.005 h^{-2} that was increased stepwise to 0.008 h^{-2} between 280 and 330 h.

The objective of this section is to demonstrate the extrapolation capability of the dynamic model, whereas a detailed discussion of the methods and results may be found elsewhere (manuscript in preparation, first author: K. Pellicer Alborch).

The model predicted the trend of all state variables during the accelerostat (Fig. 2.6). Qualitative predictions were obtained for all state variables while the biomass concentration (Fig. 2.6 B) could be quantitatively predicted. Fluctuations of the concentrations that were measured for the state variables could not be predicted. For example, the lactic acid concentration increased between the dilution rate of 0.2 and 0.4 h^{-1} (Fig. 2.6 C). At the same time, the galactose concentration was decreased (Fig. 2.6 D). It seemed that the carbon flux was shifted from galactose secretion to the production of lactic acid during this period, which could not be captured by the model. The washout of the cells was overpredicted (predicted later than measured) with an error of $D = 0.1$ h^{-1}.

Even though the model was trained using batch cultivations, it was capable of predicting the continuous accelerostat cultivation in a sufficient manner. For this reason, the model was also applied to predict the optimal conditions of a continuous large-scale production process in **Paper D**. In

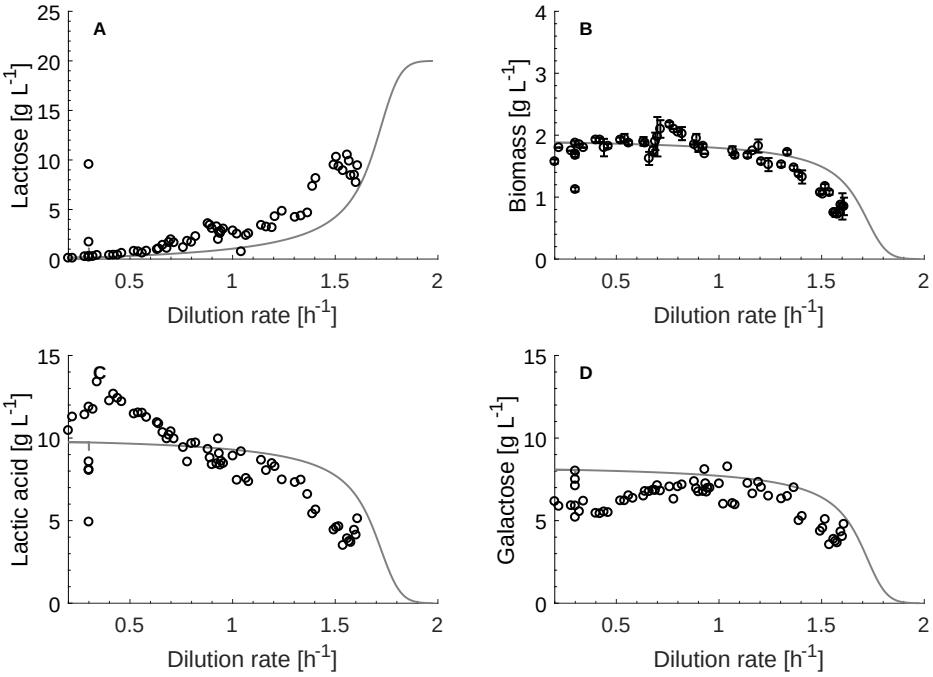


Figure 2.6: Increasing dilution rate dependent *S. thermophilus* metabolism characterized in an accelerostat cultivation. Lactose (A), biomass with standard deviation (B), lactic acid (C), and galactose (D) concentrations.

Paper D, the concentration of the carbon source in the feed and the feed flow rate were optimized to obtain the maximum cell concentration in the bioreactor effluent while minimizing the wastage of unused carbon.

All in all, the results of the validation step indicated the validity of the model. The model was therefore applied for further applications in this study as described in the following chapters. Nevertheless, a more comprehensive design of experiments covering further cultivation conditions should be performed in the future. For example, to determine the optimum of the pH function more accurately, further pH conditions, like pH = 6.1, 6.2, 6.3, 6.4, etc. could be investigated – with replicates, of course. Furthermore, the structure of the model could be further investigated and extended. The implementation of a temperature-dependent function would expand the flexibility of the model. Experiments with dynamic changes, like a pH that is changing dynamically, for example, with a sinus curve shape, could improve the predictability of the model. In addition, experiments with spikes of lactose

and lactic acid and other acids, like phosphoric acid, could be performed in order to increase the information that can be obtained in parameter estimation seeing that correlations between the carbon source, biomass, and lactic acid concentrations would be broken.

Chapter 3

Probabilistic On-line Monitoring Using a Soft Sensor

A dynamic model has been developed and validated in the previous chapter. In this chapter, the model is applied as a soft sensor to monitor lactic acid bacteria cultivations at the laboratory scale. Firstly, the framework of the soft sensor is described, which includes a Monte Carlo simulation to account for uncertainties in model prediction. The probabilistic soft sensor was then tested on three historical data sets.

This chapter is based upon the following article:

Paper B: Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. Biochemical Engineering Journal, 135, 49–60.

3.1 Framework of the Soft Sensor for On-line Monitoring

The objective of the soft sensor was to monitor the unmeasured and measured state variables of the *S. thermophilus* cultivation. It predicted the biomass, lactose, and lactic acid concentrations, as well as the pH and quantity of added ammonia among others. Both the current state of the cultivation as well as the future course of the cultivation were forecasted

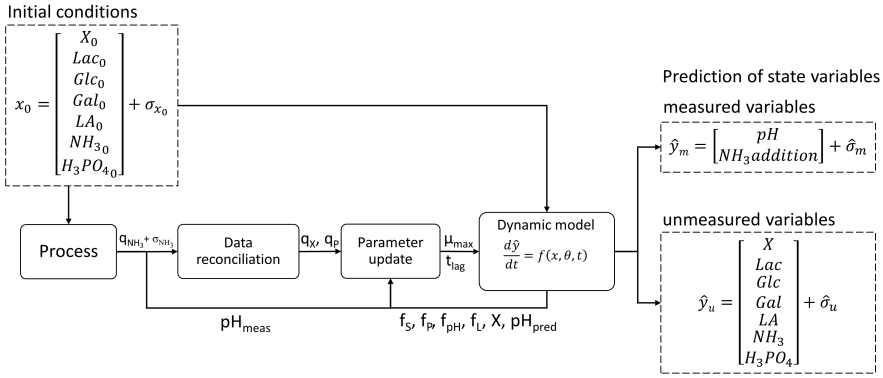


Figure 3.1: Block diagram of the soft sensor. The initial conditions for the model were defined as specified for the process. While the cultivation process is running, the on-line measured ammonia addition rate q_{NH3} was used as input for the data reconciliation module to update the biomass growth (q_X) and lactic acid production (q_P) rates. The parameter update module used the updated rates and the pH as input to update the model parameters μ_{max} and t_{lag}. Monte Carlo simulations of the dynamic model were performed considering uncertainties in the initial lactose and biomass concentrations, measured ammonia addition, and model parameters. Reprinted from **Paper B**.

by the soft sensor. The soft sensor was updated in five minute intervals. To this end, the model parameters, μ_{max} and t_{lag}, were updated using the limited available on-line measurements, namely the pH and quantity of added ammonia solution. The parameters were updated in two steps – the data reconciliation and parameter update step and then used as the inputs to the dynamic model (Fig. 3.1). To ensure robust model prediction, a Monte Carlo simulation of the model was performed at every interval, therefore accounting for uncertainties in model parameters, initial process conditions, and measurement of the ammonia addition. The model therefore predicted a probability distribution of the state variables.

Data Reconciliation Step

In the data reconciliation step, the quantity of added ammonia that was measured in real time was used to calculate the current biomass growth and lactic acid production rate. First, the lactic acid production rate (q_P) was calculated with the charge balance (Eq. (3.1)) assuming that the added

ammonia counteracts the produced lactic acid, and that the sum of charges of the other components remained constant. Second, the biomass growth rate (q_X) was calculated with the lactic acid production rate expression (Eq. (3.2)). The updated q_X was then employed during the parameter update step.

$$NH_4^+ + C_3H_5O_3^- = q_{NH,add} + q_P = 0 \tag{3.1}$$

$$q_X = \frac{q_P}{\alpha} \tag{3.2}$$

Parameter Update Step

In the parameter update step, the model parameters, μ_{max} and t_{lag}, were updated. μ_{max} was updated with the biomass growth rate expression in an iterative procedure (Eq. (3.3)). The current function values (f_{lag},f_S,f_P, and f_{pH}) were required in this procedure, and were obtained from a new evaluation of the dynamic model. As the functions were dependent on the model that itself is a function of μ_{max} among others, the iterative procedure was carried out until the change of μ_{max} was smaller than 5 % compared to the previous iteration.

q_X could not be directly employed as an input to the model because it would have been only true for the current interval. However, the soft sensor was intended to predict the future of the cultivation as well, and limitation and inhibition effects change q_X dynamically.

The lag-time parameter, t_{lag}, was updated based on pH measurements. The studied cultivations began at a pH that was higher than the control set point typically. t_{lag} was therefore adjusted during an iterative procedure so that the measured and predicted pH corresponded (Eq. (3.4)).

$$\mu_{max,k} = \frac{q_{X,updated}}{f_{lag,k-1} \cdot f_{S,k-1} \cdot f_{P,k-1} \cdot f_{pH,k-1} \cdot X_{k-1}} \tag{3.3}$$

$$t_{lag,k} = t_{lag,k-1} + (t_{pH,measured} - t_{pH,predicted}) \tag{3.4}$$

t_{lag} was updated in five min intervals until the pH control set point was reached. As long as t_{lag} was updated, the model was run from beginning to end. Once the control set point was fixed and t_{lag} was no longer updated, the current state of the prediction was saved and used as initial conditions for the next iteration.

Monte Carlo Simulation of the Dynamic Process Model

The Monte Carlo method included three steps as described in detail in **Paper B**:

Step 1: The input uncertainties were identified and defined. In this work, uncertainties in the model parameters, the initial conditions, and the on-line measurements were considered.

Step 2: N independent input sample sets (N = number of sample sets) were generated using the LHS technique (McKay et al., 1979; Sin et al., 2009b).

Step 3: The Monte Carlo simulation was then performed. Each input sample was used in the dynamic model and hence the simulations predicted the outputs N times. The probability distribution of the model outputs was later utilized in **Chapter 5** for risk quantification.

3.2 Monitoring of Lab-scale Cultivations

The probabilistic soft sensor was applied to data sets of three historical cultivations as the experimental setup did not allow accessing the on-line measurements in real time, e.g., from an OPC server (Figs. 3.2 - 3.4). Nevertheless, the actual on-line measurements, namely the pH and quantity of added ammonia solution, were used in these demonstration tests as they would have been available on-line.

The results will be explained for one cultivation shown in Fig. 3.2. Figs. 3.3 and 3.4 depict the results for the additional cultivations that substantiated the quality of the soft sensor. Furthermore, there are movies in the supplementary material of **Paper B** that show the virtual implementation

of the soft sensor. There, the representation that was intended for a plant operator and process engineer are presented as the soft sensor was updated in five min intervals. The figures, however, portray the predictions of the soft sensor at selected time points only (in the rows).

The cultivation began at an initial pH 7 (Fig. 3.2). Until the pH dropped to the control set point, ammonia solution was not added to the bioreactor, hence the only valid on-line measurement was the pH. The pH information was used to update the lag-time parameter. For demonstration purposes, the lag-time parameter was not updated in this example at 2 h (Fig. 3.2, first row), and the initial value for $t_{lag} = 1$ h was used. The soft sensor predicted both the measured state variables (pH and ammonia in the left column) and unmeasured state variables (biomass in the middle column and lactose and total lactic acid in the right column). For comparison, the off-line measurements are shown here as well, even though they would not be available in an on-line implementation. After 2 h, there is a clear offset between the predictions and actual values as t_{lag} had not been updated yet.

The Monte Carlo simulation propagated the input uncertainties to the outputs. The 95 % confidence intervals of the model predictions for the biomass, lactose, and lactic acid concentrations are shown (Fig. 3.2). They allow for a robust interpretation of the results.

After 2 h and 40 min (Fig. 3.2, second row), the pH control set point was arrived at. t_{lag} was finally updated, and there was already a solid prediction of the state variables. From this point forward, the addition of ammonia solution started in order to control the pH, and μ_{max} was updated accordingly.

The procedure to update t_{lag} introduced a dependency of the soft sensor on pH measurement. In the case pH measurement was faulty, the soft sensor will not recognize it. Nevertheless, given the limited number of on-line measurements, this might be the only possibility for adjusting the model with respect to the lag time that naturally varies from batch to batch. It is therefore advised to use at least two independent pH probes to double verify pH measurement.

After 3 h (Fig. 3.2, third row), the ammonia had been added already but

the controller overshot so that currently, just little ammonia was added. μ_{max} was therefore updated to a low value, and the soft sensor predicted almost no growth for the coming hours. The soft sensor prediction was therefore incorrect at that moment. However, this mechanism could also be used to warn the operators in case the base addition would stop for a long period of time. In case no base is added, two scenarios are possible: 1) lactic acid production could have stopped. In the case no lactic acid was produced, it is very likely that the growth stopped as well; and 2) there was a failure in the base addition unit. This could be that of a pump or IT component among others. Depending on the pH profile and, e.g., pump signals, the error could be narrowed down. Fast action of the process operators would then be needed in order to solve the problem.

As the cultivation progressed, more and more on-line data was available. μ_{max} was updated every five min according to the base addition measurements, and a reliable prediction of the unmeasured states was achieved already after 4 h (Fig. 3.2, fourth row) and later after 6 h (Fig. 3.2, bottom row).

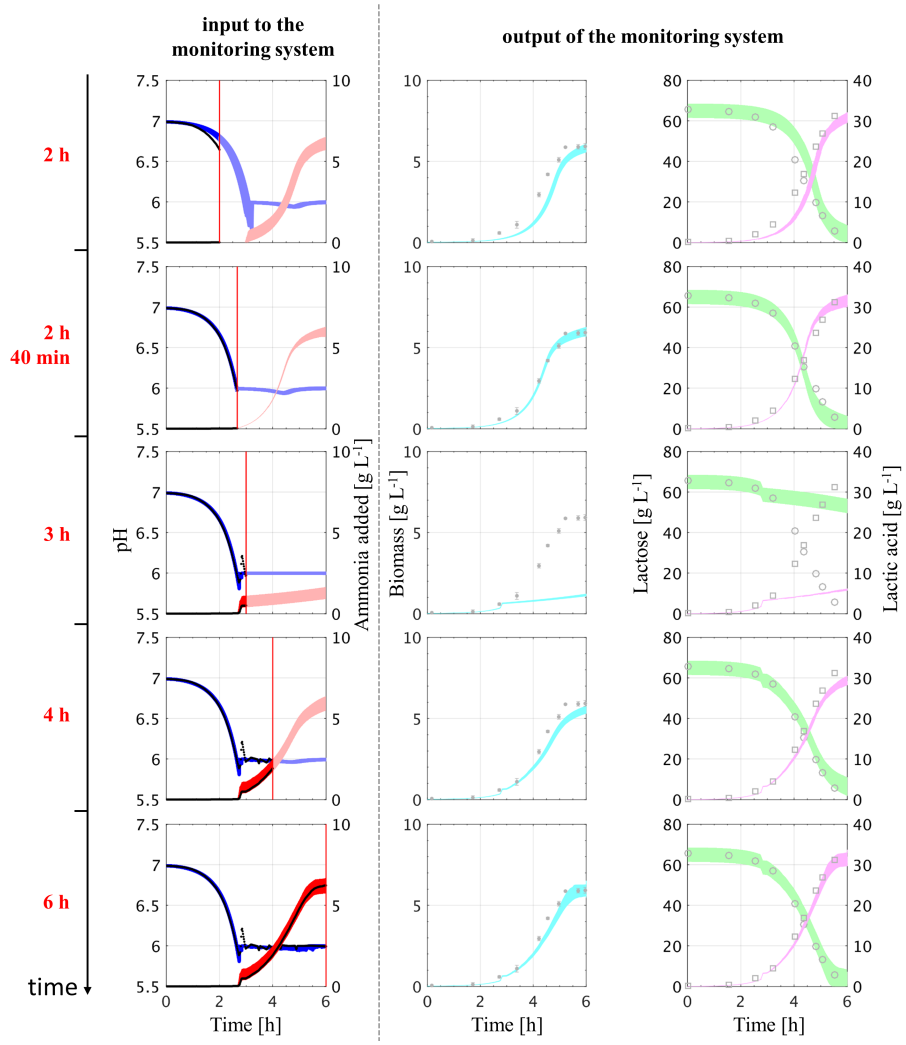


Figure 3.2: The probabilistic monitoring system applied to the lab-scale batch data of a *S. thermophilus* cultivation (1st demonstration data set). The monitoring system reads in the on-line available data (left column, black dots), ammonia addition quantity, and pH, and predicts the state variables (middle and right column) every five min. A total of 100 Monte Carlo simulations of the dynamic model were performed. The 95 % confidence intervals of the probabilistic model predictions are shown at five time points during the cultivation (2 h, 2 h 40 min, 3 h, 4 h, 6 h). Predictions of the pH (blue), ammonia addition (red), biomass (cyan), lactose (green), and lactic acid (magenta) concentrations are shown. The off-line measurements for biomass (grey dot with standard deviation), lactose (grey circle), and lactic acid (grey square) are shown for comparison only, but were not used to update the soft sensor. Reprinted from **Paper B**.

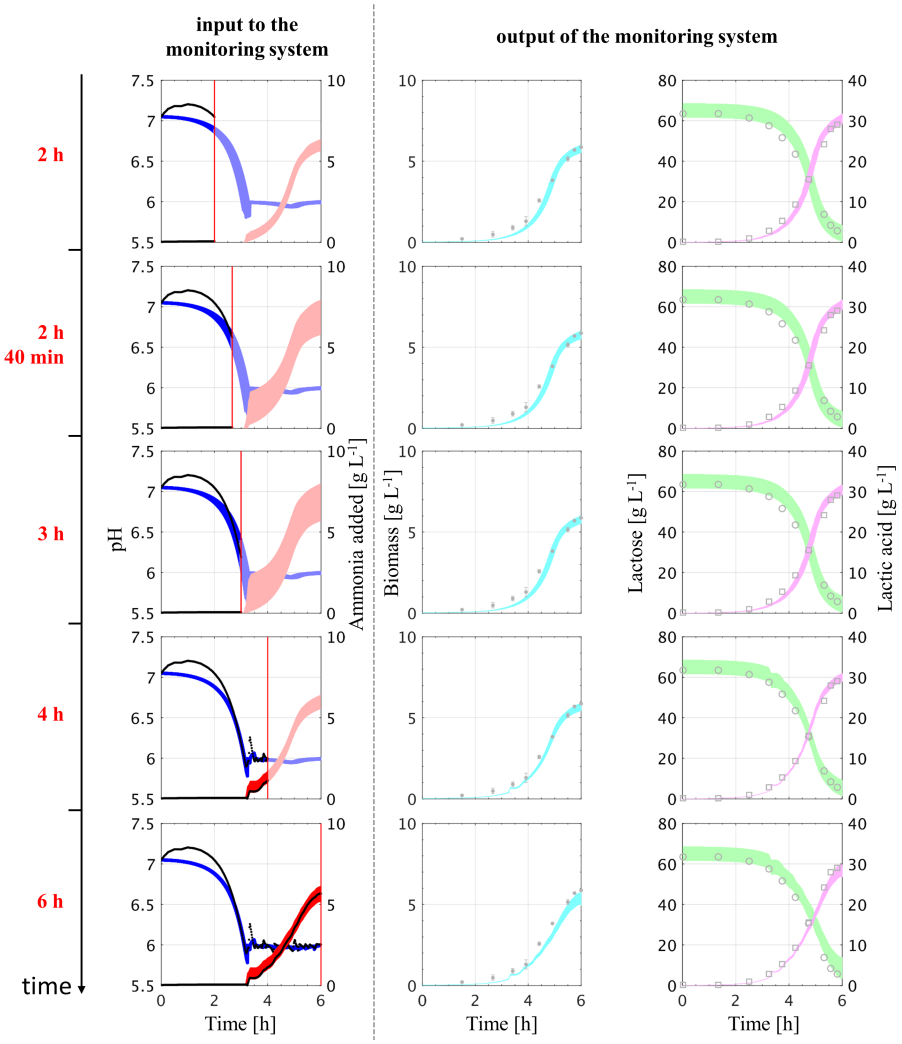


Figure 3.3: The probabilistic monitoring system applied to the lab-scale batch data of a *S. thermophilus* cultivation (2nd demonstration data set). Reprinted from **Paper B**.

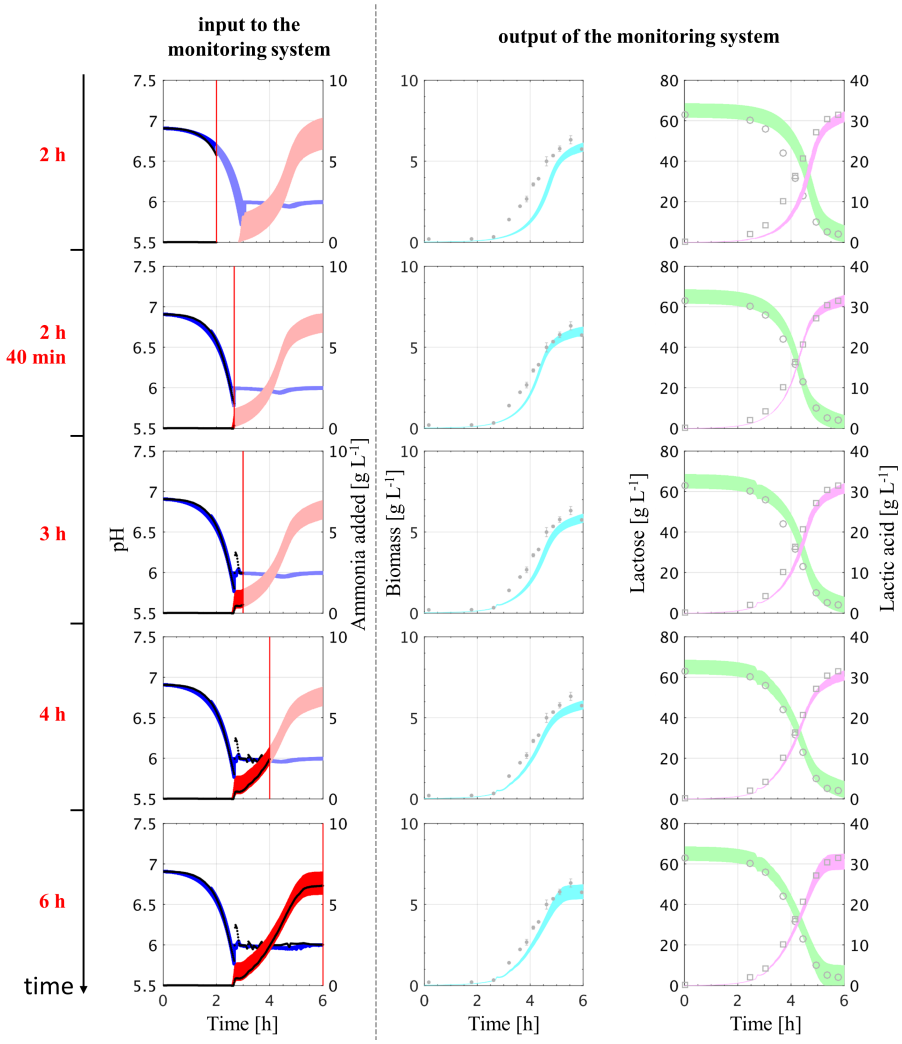


Figure 3.4: The probabilistic monitoring system applied to the lab-scale batch data of a *S. thermophilus* cultivation (3rd demonstration data set). Reprinted from **Paper B**.

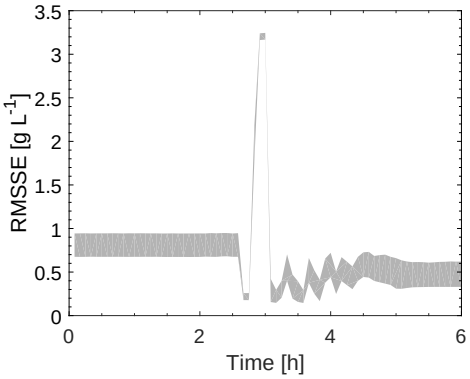


Figure 3.5: Error of the biomass prediction in the soft sensor; the RMSSE with the 95 % confidence interval for the biomass prediction in the first demonstration data set (see Fig. 3.2). Reprinted from **Paper B**.

To measure the quality of the model predictions, the RMSSE of the cultivation in Fig. 3.2 was assessed. At the beginning of the cultivation, the RMSSE for biomass was 0.8 $\pm$ 0.1 g L^{-1} (Fig. 3.5). After 3.5 h, the RMSSE remained lower than 0.6 g L^{-1}. At the end of the cultivation, the RMSSE was 0.5 $\pm$0.1 g L^{-1}, which corresponded to an error of less than 10 % with respect to the final total biomass concentration. As discussed before, the accuracy of the soft sensor was the lowest when too much ammonia solution was added and the pH overshot. The ammonia addition stopped subsequently (after ca. 3 h), hence μ_{max} was underpredicted, and the RMSSE was larger than 3 g L^{-1}.

The reproducibility of the soft sensor predictions was tested with 10 different input sample matrices for the Monte Carlo simulation. The RMSSE of biomass varied less than 0.5 %. Furthermore, the tolerance limit in the iterative update procedure for μ_{max} (Eq. (3.3)) was evaluated. Changing the limit from 5 % to 1 % and 0.1 % did not influence the prediction accuracy in this work. Nevertheless, this might be necessary for other applications.

There are several studies that have implemented either a model-based or data-driven soft sensor to monitor lactic acid bacteria cultivations. Acuña et al. (1994) and Peter and Röck (2012) applied a soft sensor that used the pH measurement and base addition information as well. However, the latter application is limited to the prediction of the lactic acid concentration. A data-driven soft sensor was proposed by Fayolle et al. (1997) and Payot et al.

(1997). Fayolle et al. (1997) employed mid-infrared spectroscopy while Payot et al. (1997) utilized the conductivity signal. In general, the conductivity signal might be a valuable measurement that could be included in the soft sensor in the future. It could be especially useful for fault detection as it is correlated with lactic acid concentration, and could be employed to validate the pH and base addition measurements. However, for the main model, a model-based framework was opted for over a data-driven approach. Firstly, the advantage of a model-based over a data-driven approach is that the cause of poor model predictions could be found, understood, and corrected. As an example, one could change the algorithm not to update μ_{max} directly after the control set point is reached. In this way, the faulty prediction after 3 h (Fig. 3.2, time = 3 h) could be corrected. These modifications would be more difficult with a data-driven model. A further advantage of the model-based approach is that it could be easily adapted to other cultivations, e.g., other homolactic lactic acid bacteria strains, by readjusting a number of the kinetic parameters. In addition, it might be also applicable to other cultivation conditions as already outlined with the A-stat cultivation in **Chapter 2** without the need to re-calibrate the model as it would be necessary with a data-driven approach.

In summary, the probabilistic model-based soft sensor was validated with 2-L lab-scale experiments. In contrast to earlier studies, this work considered several sources of uncertainties using the Monte Carlo procedure, and demonstrated a probabilistic monitoring system. However, further uncertainties, for example in the chemical model, and stochastic uncertainties, like pump failures, could be included in future investigations. Furthermore, if the soft sensor would be applied at the lab-scale on-line, different scenarios to test the detection of faulty batches by the soft sensor should be performed. These experiments could include a deliberate stop of the base addition simulating a pump failure, or the addition of, e.g., ethanol to stop biomass growth. In **Chapter 5**, the soft sensor will be applied to a 700-L cultivation, and the probabilistic model outputs will then be used to quantify the risk of not producing the target biomass yield.

Chapter 4

Prediction of pH Gradients

Chapter 3 described the application of the kinetic model as a soft sensor for on-line monitoring of lab-scale cultivations. At a large scale, however, gradients emerge and often have an influence on the metabolic activity of the cells. In this work, pH gradients were considered as the key parameter of lactic acid bacteria batch cultivations. In order to apply the model at a large scale, fluid dynamics need to be taken into account. In this chapter, firstly, a CFD model was employed to predict pH gradients during a cultivation in a 700-L bioreactor and validated with experimental measurements. Secondly, a CFD-based compartment model was applied, and then benchmarked against the experiments and CFD predictions.

This chapter is based upon the following articles:

Paper E: Spann, R., Glibstrup, J., Pellicer Alborch, K., Junne, S., Neubauer, P., Roca, C., Kold, D., Eliasson Lantz, A., Sin, G., Gernaey, K. V., & Krühne, U. (2018). CFD predicted pH gradients in lactic acid bacteria cultivations. Biotechnology and Bioengineering. **In preparation**

Paper F: Spann, R., Roca, C., Gernaey, K. V., & Sin, G. (2018). A Validated CFD-based Compartment Model to Assess pH Gradients in Lactic Acid Bacteria Cultivations. AIChE Journal. **Submitted**

4.1 Computational Fluid Dynamics (CFD) Model

The objective of this chapter was to consider the fluid dynamics of a large-scale bioreactor, in particular to predict pH gradients during the *S. thermophilus* cultivation. The pH gradients are important for assessing the mixing of the bioreactor seeing that pH gradients could lead to a productivity loss of the cells.

As a case study, a stirred-tank bioreactor with a liquid volume of 700 L was made use of. Firstly, a CFD model was developed and then validated using tracer-pulse simulations (Fig. 4.1). The aforementioned biokinetic model was eventually integrated into the CFD simulation and a *S. thermophilus* cultivation, including its pH gradient, was simulated. The results were validated with one experiment. Secondly, a compartment model was designed to speed up simulation time. The compartment model was based on the velocity profile that was obtained from the CFD simulation. It was also applied to predict the cultivation and pH gradient while being benchmarked against the CFD prediction.

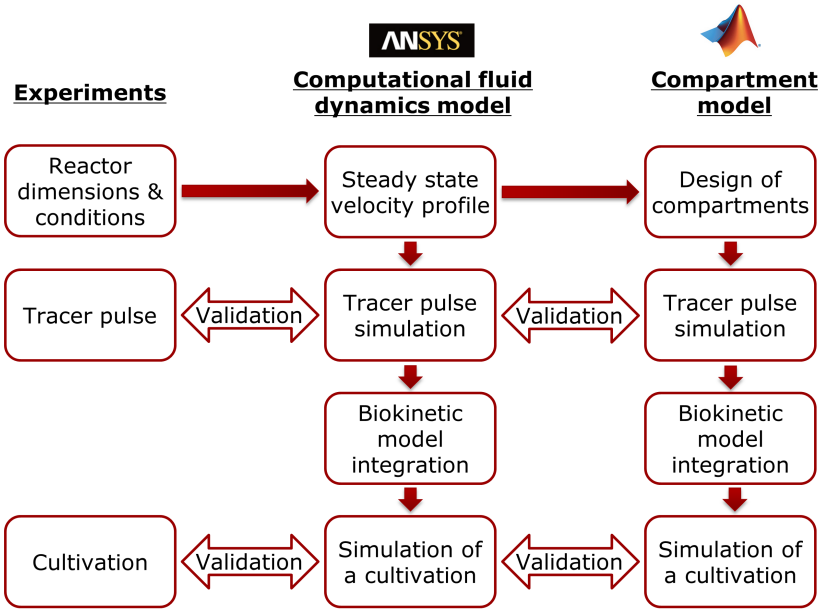


Figure 4.1: Framework for the CFD and compartment model simulations.

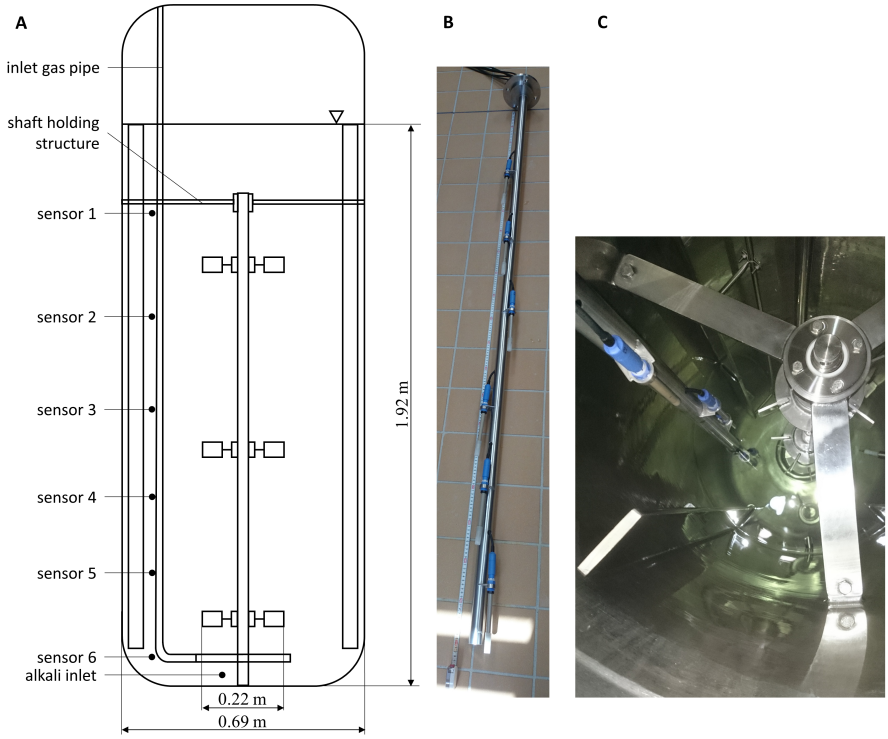


Figure 4.2: Bioreactor set up for the mixing time experiments and cultivation. A: Simplified geometry of the stirred-tank bioreactor. B: Lance holding the six pH sensors. C: Top view into the bioreactor with the pH sensor lance. Adapted from **Paper E.**

The bioreactor was equipped with three six-blade Rushton turbines and four baffles (Fig. 4.2). It was utilized for the CFD and compartment model studies (**present chapter**) as well as the risk quantification (**Chapter 5**). Six pH sensors were placed at different vertical positions in the bioreactor and they recorded the pH every second during the tracer-pulse experiments and cultivation. The bioreactor was operated at 130 rpm during the cultivation and 240 rpm for the tracer-pulse experiments.

A simplified bioreactor geometry was designed in SolidWorks (Dassault Systèmes, Vélizy-Villacoublay, France) as the gas inlet pipe, including the sparger ring, and the shaft-holding structure were omitted. Half of the volume was modelled applying a rotational periodicity plane. The bioreactor consisted of three rotating domains for the impellers along with a stationary

domain. Structured meshes with six-sided hexahedral elements were defined in in ANSYS ICEM CFD 17.1 (ANSYS, Inc., PA). The entire mesh consisted of ca. 1.6 million nodes. The simulations were performed in ANSYS CFX 17.1, and the k-ε turbulence model was applied to model fluid dynamics. As the cultivation was not aerated, a one-phase CFD model could be applied. Details on the simulation settings could be found in **Paper E**.

First, a steady-state velocity profile was obtained (Fig. 4.3). The CFD simulation predicted that six recirculation loops were generated by the three Rushton turbines as it was also expected based on earlier studies with a similar configuration (Vrabel et al., 2000). It is predicted that mixing is fast within the recirculation loops while the liquid flow between compartments is slow. The pH sensors were therefore placed in the predicted recirculation loops during the subsequent experiments (the tracer pulses and cultivation) (Fig. 4.2 B and C). The cross-section from the top (Fig. 4.3 A-A) depicts high circumferential velocities close to the impeller blades that turn with four rounds per minute (240 rpm). Close to the baffles, the circumferential mixing is much slower.

CFD Model Validation with Mixing Experiments

The CFD model was then validated with tracer-pulse experiments. For the experiments, the bioreactor was filled with 700 L of tap water (35 °C) and concentrated NaOH solution was added as tracer from the top of the bioreactor. The distribution of the tracer was modelled in a dynamic CFD simulation while one mol of an additional variable was specified in a cylindrical volume at the top of the bioreactor. The dynamic simulation was performed with a transient (time-dependent) velocity profile using a steady-state result as the initialization state.

The tracer-pulse simulation revealed the fast mixing within the recirculation loops (fast radial and circumferential mixing) but slow axial mixing (Fig. 4.4, and a movie in the Supplementary Material of **Paper E**). The tracer required several seconds to pass from one recirculation loop to another.

The dynamic response of pH in the mixing experiments and the concentration of the tracer in the CFD simulation were monitored at different positions

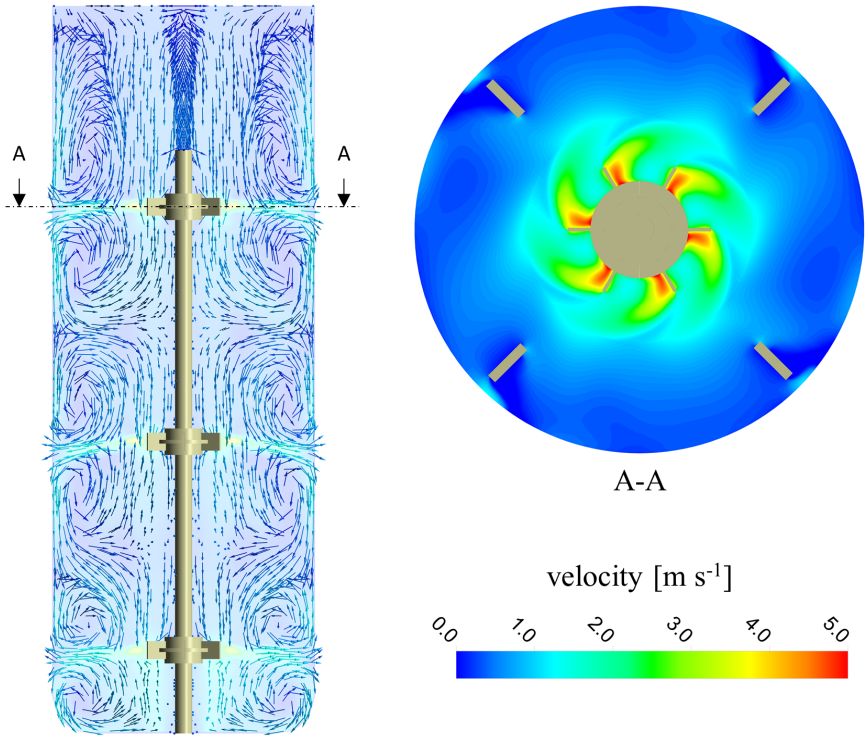


Figure 4.3: Steady-state solution of the 700-L bioreactor at 300 rpm. Left: velocity streamlines with velocity in stationary frame. Right: contour plot of the circumferential velocity in stationary frame. Reprinted from **Paper E.**

to understand the mixing dynamics when base is added in the cultivation to control pH (please note that the base was actually added at the bottom of the bioreactor in the cultivation but for the mixing experiments, base addition from the bottom was not possible owing to technical issues). At the top sensor location 2, the pH overshot first and then reached a stable value (Fig. 4.5 A). It is unfortunate that the very top sensor 1 failed to track the data during the experiment. The pH sensors 3 - 6, which were placed further away from the base addition position, responded with a sigmoidal curve. In order to compare the pH measurements with the CFD predictions, the measurements were normalized:

$$pH_i'(t) = \frac{pH_i(t) - pH_i(t=0)}{pH_i(t=\infty) - pH_i(t=0)} \tag{4.1}$$

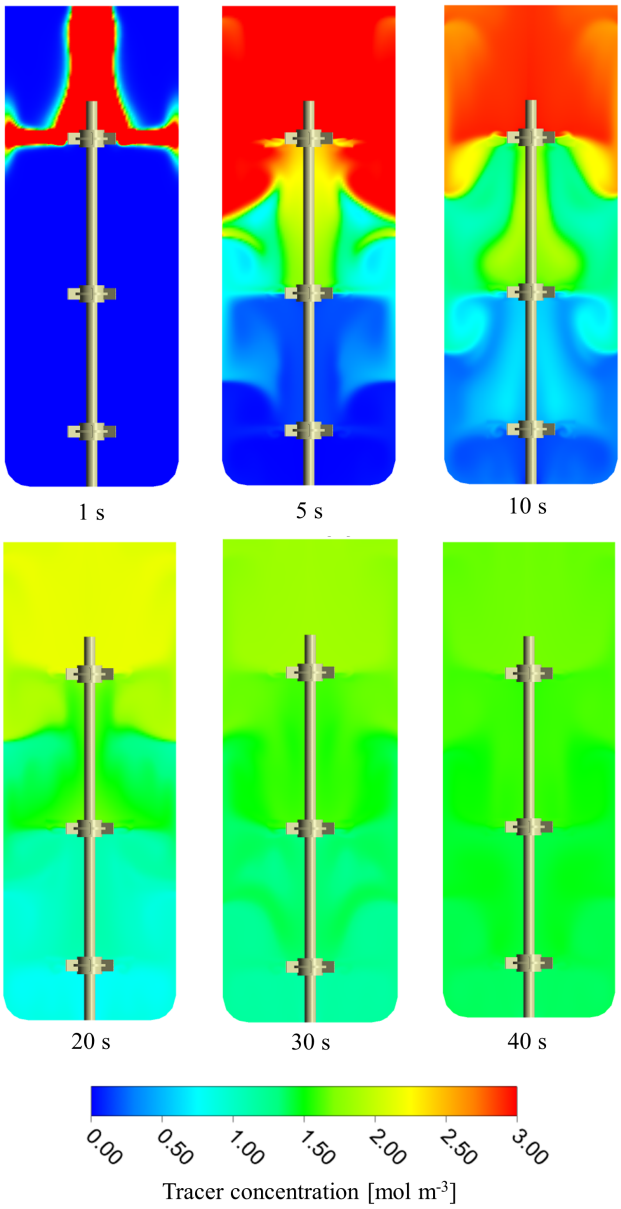


Figure 4.4: Distribution of the tracer in the tracer-pulse simulation. A dynamic simulation with a transient velocity profile at 240 rpm was performed. Reprinted from **Paper E.**

where pH_i' is the normalized output of the ith sensor/monitoring position. For t=∞, the average measurements between 4.5 and 5 min after each pulse

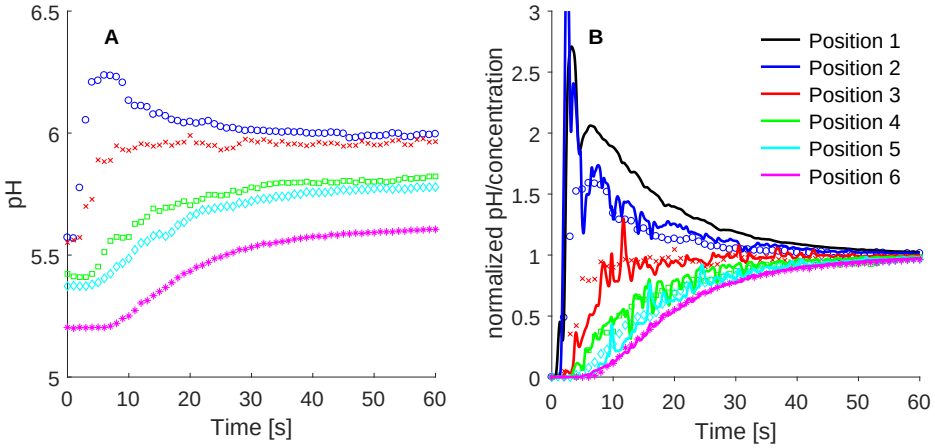


Figure 4.5: Dynamic response at the six measurement and monitoring positions after the tracer pulse. A: Measured pH in the experiments. B: Normalized values of the experiments (symbols) and CFD simulation (lines). Adapted from **Paper E.**

were considered. The concentrations of the CFD simulation were normalized accordingly.

The simulated dynamic response at the different sensor locations matched the measurements quantitatively (Fig. 4.5 B). Both the dynamic shape and order of magnitude could be predicted accurately. However, minor oscillations that were predicted by the CFD simulation were not measured. This could be attributed to the response time of the ISFET pH sensor. The sensors employed needed 4 - 8 s to reach the final pH value (± 0.02) under well-mixed conditions (data not shown). They might therefore by sluggish to capture rapid changes.

The mixing time was calculated based on the normalized measurements and predictions at the different positions. To this end, the logarithmic squared deviation with respect to the normalized upper bound 1 was calculated according to Paul et al. (2003):

$$log\, D^2 = log\left[\frac{1}{n}\cdot\sum_{i=1}^{n}\left(pH_i'(t)-1\right)^2\right] \tag{4.2}$$

where n represents the number of sensors. $log\, D^2 = -2.65$ when 95 % homogeneity is achieved.

Table 4.1: Mixing times at 240 rpm to reach 95 % homogeneity.

Experiments [s]	CFD simulation with a transient velocity profile [s]
42	
50	46
51	

Note. Adapted from **Paper E**.

A mixing time of 48 $\pm$ 5 s was measured in the experiments (three replicates) to achieve 95 % homogeneity (Table 4.1). This was congruous with both the predicted 46 s of the CFD simulation and reported mixing times of similar bioreactors in the literature (Delvigne et al., 2006). At the production scale, mixing times on the same order of magnitude were assessed (internal reports). It was therefore concluded that the present case study with the 700-L bioreactor would offer representative insight into gradients in production-scale processes.

Both the prediction of the dynamic response at the sensor positions and the mixing time supplied convincing evidence that the CFD model described the fluid dynamics of the bioreactor at 240 rpm satisfactorily.

Prediction of pH Gradients in a 700-L Cultivation

In order to simulate the *S. thermophilus* batch cultivation, the biokinetic model was integrated in the CFD simulation. The mixed weak acid/base model that was applied with the purpose of predicting pH in **Chapter 2** and **Chapter 3** could not be solved in CFX as it resulted in a stiff system with fast (mixed weak acid/base model) and slow (biokinetic model) differential equations. An algebraic linear equation (Eq. (4.3)) was used instead. The equation was derived from data of lab-scale experiments.

$$pH = -0.44 \cdot (C_P - 5.29 \cdot C_{NH_3}) + 7.00 \tag{4.3}$$

In the cultivation, the pH was controlled by adding NH_4OH at the bottom of the bioreactor. The controlling pH sensor was located close to sensor 5.

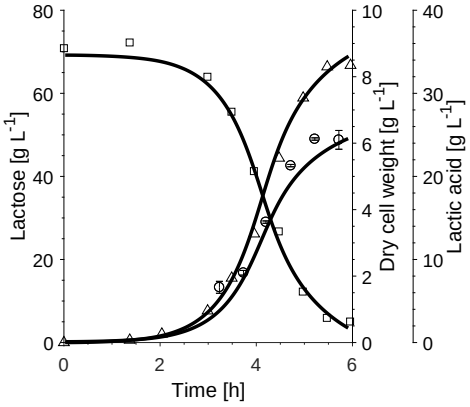


Figure 4.6: CFD predictions (lines) and measurements of the lactose (squares), biomass (circles), and lactic acid (triangles) concentrations in the 700-L cultivation. Reprinted from **Paper E.**

The CFD simulation was set up accordingly with a P-controller with the step function:

$$NH_{3,add} = step(6 - pH) \cdot (6 - pH) \cdot 11900 \cdot g \cdot h^{-1} \tag{4.4}$$

The cultivation was conducted at a stirrer speed of 130 rpm. Unfortunately, there were no mixing experiments available to validate the CFD model under these conditions. However, the CFD model was assumed to predict the fluid dynamics under these conditions properly, as well. A transient (time-dependent velocity field) tracer-pulse simulation at 130 rpm predicted a mixing time of 85 s to reach 95 % homogeneity. In order to simulate the cultivation in a considerable time period, a steady-state velocity profile was required for the simulation. However, the mixing times predicted with a transient and steady-state velocity profile did not match. A steady-state velocity profile with a stirrer speed of 200 rpm instead predicted a mixing time of 89 s, and was therefore deemed to mimic the fluid dynamics at the real speed of 130 rpm sufficiently.

The dynamic CFD simulation (with the steady-state velocity profile) of the *S. thermophilus* cultivation predicted the measured biomass growth, substrate consumption, and lactic acid production accurately (Fig. 4.6). After 6 h of cultivation, 6 g L^{-1} biomass and 34 g L^{-1} lactic acid were yielded.

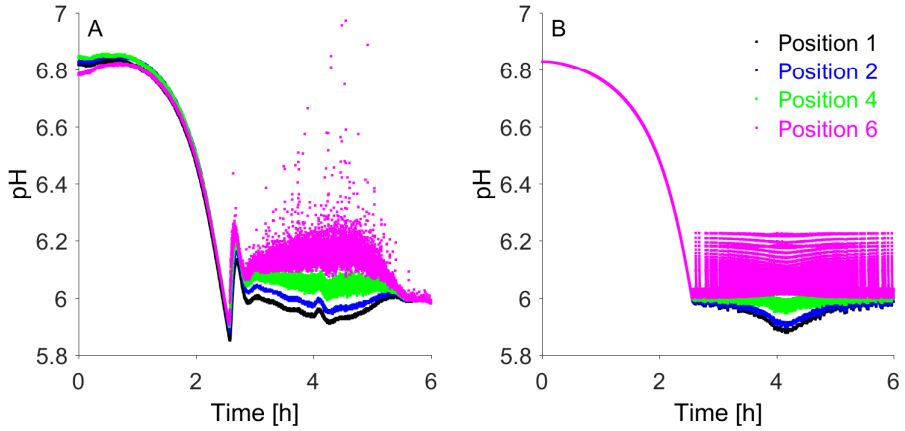


Figure 4.7: pH at different vertical locations during the 700-L cultivation. A: pH measurements at position 1, the top of the bioreactor (black), 2 (blue), 4 (green), and 6, the bottom of the bioreactor (pink). B: CFD prediction. Reprinted from **Paper E.**

The pH was also accurately predicted at the different positions (Fig. 4.7). The pH was 6.8 at the beginning of the cultivation and dropped to the controlling pH value $pH_{set} = 6$ within 2.5 h. Then, the pH was controlled by adding the base below the bottom impeller (Fig. 4.2). The measurements (Fig. 4.7 A) indicated a maximum pH gradient between 5.9 at the top of the bioreactor and 6.2 at the bottom sensor of the bioreactor at ca. 4 h. A few measurements, which were recorded every second, were as high as pH = 7. The CFD simulation (Fig. 4.7 B) predicted similar values with deviations of less than 0.05 pH units compared to the measurements.

The multi-position pH measurements suggested that pH gradients existed in the studied 700-L bioreactor. Furthermore, the pH gradients could be quantitatively predicted with the CFD simulation. However, the dynamic CFD simulation of the cultivation required four days of computation time on 20 CPU cores within the DTU High Performance Computing Cluster (https://www.hpc.dtu.dk/).

4.2 Compartment Model

The compartment model was therefore developed with the aim of speeding up the simulation. This would allow applying the model for on-line applications, such as the soft sensor for monitoring, where computation time is crucial.

The compartment model was designed based on the steady-state velocity profile that was obtained from the CFD simulation (Fig. 4.1). In alignment with the methodology that was applied for the CFD model, the compartment model was validated with the tracer pulse, and subsequently coupled with the biokinetic model to simulate the cultivation.

Compartment Model Development

In this work, a compartment model consisting of seven compartments was designed based on the CFD simulation. Each recirculation loop was represented by a compartment (Fig. 4.8). The location of the compartment boundaries and the flows between the compartments were calculated as described in detail in **Paper F**. The flows were exported from the CFD software, and the compartments were separated where the axial flows were lowest. In addition to the six compartments resembling the six recirculation loops, a seventh compartment was placed for the top 10 cm of the bioreactor. This was necessary to perform the simulations of the tracer pulse. It is noted that the applied compartment model is a simple and coarse model. However, the main aim of this study was to demonstrate the various possibilities with the compartment model, and not to optimize the compartment further.

Two compartment models were designed for the different stirrer speeds of 130 and 240 rpm that were investigated. To determine the properties of the compartment model, the steady-state velocity profile of the CFD simulation was required. However, as mentioned in the CFD model section, the results of the tracer-pulse simulation that used a transient velocity profile and a steady-state velocity profile differed when employing the same stirrer speed. The CFD tracer-pulse simulation with the transient velocity profile had predicted the measured mixing time and the dynamic responses of the pH sensors at the various locations accurately. It was therefore decided to manipulate the

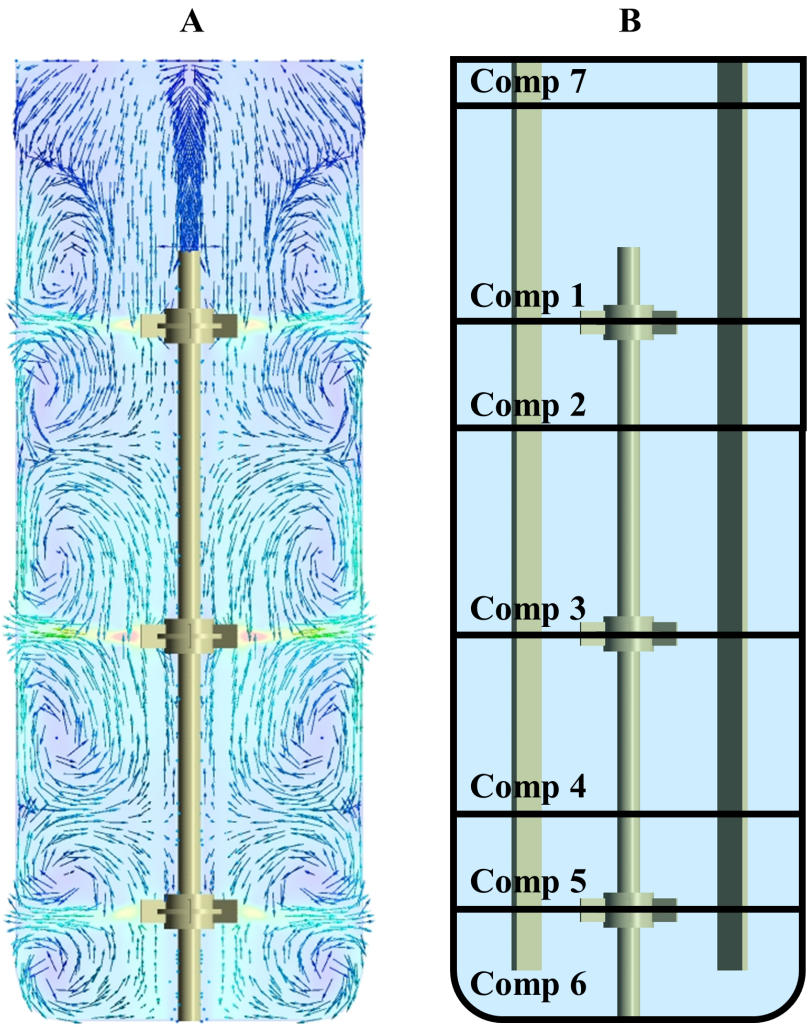


Figure 4.8: CFD-based compartment model design. A: Steady-state velocity profile at 200 rpm representing the real stirrer speed of 130 rpm that was used in the cultivation. B: Location of the seven compartments. Adapted from **Paper F.**

stirrer speed in the steady-state simulation in order to obtain the solution that matched the mixing time closely and thus the fluid dynamics of the real system. The real stirrer speed of 240 rpm that was used for the tracer-pulse experiments was mimicked with a steady-state velocity profile featuring 300 rpm. The real stirrer speed of 130 rpm that was used in the cultivation

Table 4.2: Properties of the compartment models.

Compartment number	Volume [L]		Compartment connection	Flow rates between the compartments [L s^{-1}]	
	200 rpm*	300 rpm*		200 rpm*	300 rpm*
7	19.1	19.1	7 ↔ 1	10	15.3
1	80.6	80.6	1 ↔ 2	11.7	17.5
2	39.6	55.7	2 ↔ 3	9.7	13.6
3	76.8	60.6	3 ↔ 4	14.2	18.1
4	67.1	50.9	4 ↔ 5	8.8	12.1
5	36.4	52.6	5 ↔ 6	12.2	17.9
6	38.8	38.8			

* This is the impeller speed of the steady-state velocity profile.

Note. Adapted from **Paper F**.

was mimicked with a steady-state velocity profile with 200 rpm. The design of the two compartment models was therefore based on these steady-state velocity profiles.

The properties of the compartment models are shown in Table 4.2. The volume of the compartments 6 (bottom), 1, and 7 (top) was identical for both compartment models while the volume of the other compartments differed as the predicted size of the recirculation loops was different. The flow rates between the compartments increased as expected from the compartment model for 200 rpm to the model for 300 rpm. The compartment model was implemented in MATLAB as an ODE system. For the simulation of the cultivation, it was coupled with the biokinetic and chemical model.

Compartment Model Validation with Mixing Experiments

In order to simulate the tracer-pulse addition from the top of the bioreactor with the compartment model and to assess the mixing time – as was carried out to validate the CFD model – an additional state variable was defined as the tracer. The small top compartment 7 was filled with 100 g L^{-1} of tracer initially while the tracer concentration was 0 g L^{-1} in the other compartments. The reactions of the biokinetic and chemical model were not considered for the tracer-pulse simulation.

Table 4.3: Mixing times obtained in the compartment model, CFD model, and experiments.

	Mixing time to reach 95 % homogeneity [s]	
	200 rpm*	300 rpm*
Compartment model	68	52
CFD model (steady-state)	89	55
Experiment	not available	48 $\pm$ 5 (240 rpm)

* This is the impeller speed of the steady-state velocity profile.

Note. Adapted from **Paper F**.

The mixing times predicted by the compartment model were compared with the experiments and CFD simulation (Table 4.3). At a stirrer speed of 240 rpm, the compartment model predicted a mixing time of 52 s to achieve 95 % homogeneity, matching the experimental mixing time of 48 $\pm$ 5 s. The mixing time predicted by the CFD simulation (with the steady-state velocity profile) was 55 s for these conditions. At the cultivation conditions (130 rpm), the compartment model predicted a mixing time of 68 s, which is 24 % lower than the prediction of the CFD model (89 s). Note that the CFD simulation with a steady-state velocity profile was utilized here, while the tracer-pulse simulations in Chapter 4.1 (Table 4.1) were conducted with a transient velocity profile. The presented compartment model was deemed to describe the system sufficiently even though the fluid dynamics were not captured as comprehensively as with the CFD model. Based on the fact that the biological reaction rates are slower than the fluid dynamics, the error of the compartment model can however be accepted.

Prediction of pH Gradients in a 700-L Cultivation and the Benefit of Computational Speed

In order to simulate the pH gradients in the *S. thermophilus* cultivation, the kinetic model was solved with the compartment model. First, the linear pH model, which was used in the CFD simulation to overcome the problems of a stiff ODE system in the CFD simulation, was compared with the mixed weak/acid base model. Both models were implemented in the compartment

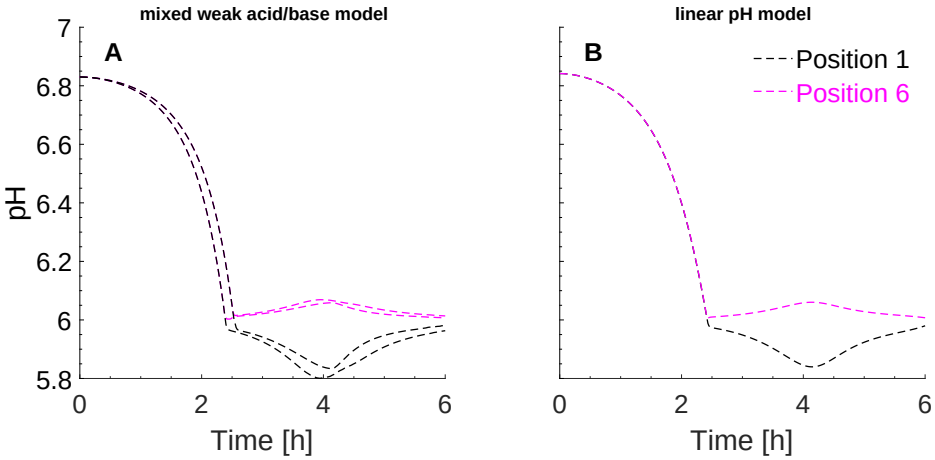


Figure 4.9: pH gradients in the *S. thermophilus* cultivation. A: The mixed weak acid/base model was applied. B: The linear pH correlation (Eq. (4.3)) was applied as in the CFD model. Adapted from **Paper F.**

model, and the simulations of the cultivation revealed marginal differences (Fig. 4.9). Both models predicted a pH gradient between 6.1 and 5.8 during the exponential growth phase. For the simulation with the mixed weak acid/base model (Fig. 4.9 A) uncertainties in the model parameters and initial process conditions were considered, and the 5 and 95 % percentiles are presented. Based on the general preference of a mechanistic model, the subsequent simulations were performed with the mixed weak acid/base model.

The compartment model predicted the pH gradients that were measured in the cultivation qualitatively (Fig. 4.10). The drop in the pH at the beginning of the cultivation was predicted accurately. During the exponential growth phase, a gradient between pH 5.9 and 6.2 was measured as discussed earlier (Fig. 4.10 A). The compartment model predicted a gradient between pH 5.8 and 6.1 in the corresponding compartments 1, 2, 4, and 6 (Fig. 4.10 B). The compartment model overpredicted the effect of the reduction of pH caused by lactic acid production in the top compartment by 0.1 pH unit. In the bottom compartment, the pH increased because of base addition. The resulting pH increase was slightly (0.1 pH unit) underpredicted by the compartment model. First of all, the deviation of the predictions and measurements on the order of 0.1 pH unit are completely acceptable as they

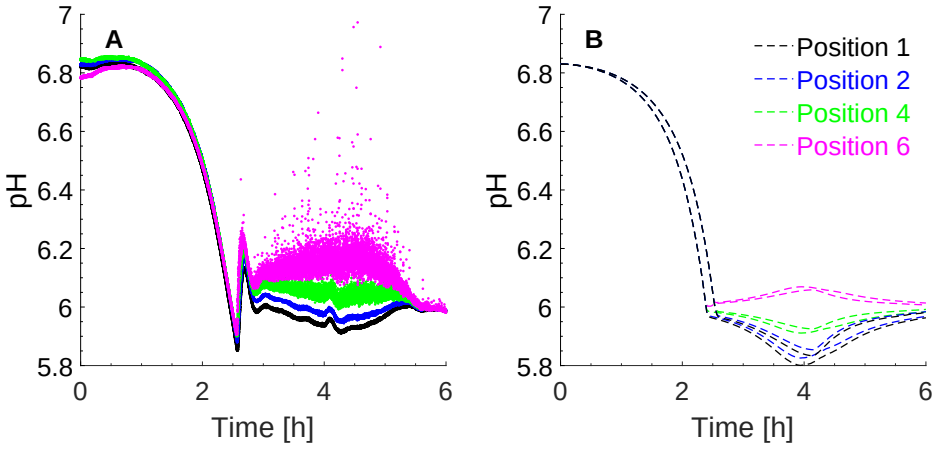


Figure 4.10: pH gradients that were measured and predicted by the compartment model in the *S. thermophilus* cultivation in the 700-L bioreactor at 130 rpm. A: pH measurements at the different vertical positions (see Fig. 4.2). B: pH predicted by the compartment model at the corresponding positions with the 95 % confidence intervals that resulted from a Monte Carlo simulation. Adapted from **Paper E and F.**

could have resulted from measurement errors and only one cultivation data set was available. Further experiments are therefore needed to evaluate the measurements statistically. However, the simple design of the compartment model could have also led to the small mismatch. A finer compartment model might therefore capture the pH at the measurement positions qualitatively in contrast to the present model that simplified each recirculation loop to one compartment and assumed the same conditions there.

The biological state variables were predicted accurately by the compartment model (Fig. 4.11), and the predictions were compared with the CFD simulation and a one-compartment simulation. The biomass (Fig. 4.11 A) and lactose concentrations (Fig. 4.11 C) were predicted similarly by the compartment model and CFD model while the yield of the total lactic acid concentration was higher in the CFD simulation than in the compartment model simulation (Fig. 4.11 B). This could be attributed to the fact that the growth-associated lactic acid production constant α was changed during this study, and a higher value ($\alpha = 5.59$ g g^{-1}) was used in the CFD simulation than in the compartment model ($\alpha = 5.19$ g g^{-1}). Furthermore, differences between the CFD and compartment models could be caused by the ODE

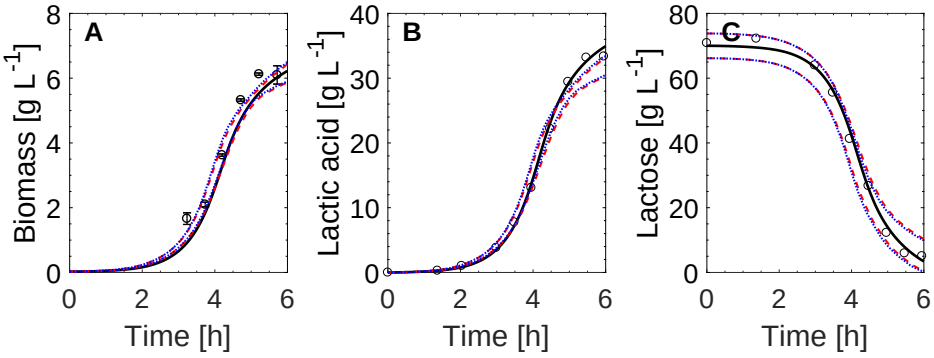


Figure 4.11: Measured and simulated biomass (A), total lactic acid (B), and lactose (C) concentrations in the 700-L cultivation. Measurements (circles), CFD simulation (black solid lines), 95 % confidence intervals of the compartment model (red dashed lines), and the homogeneous (one-compartment) model (blue dotted lines). Reprinted from **Paper F.**

solvers. A considerable impact might have had the time steps that were used in the solvers. In the CFD simulation, the time step was 1 s, while the applied ode15s solver in MATLAB made use of a dynamic time step for the compartment model simulation. There were no significant differences between the compartment model and homogeneous one-compartment cultivation (Fig. 4.11). In the case the purpose of the model is to describe the microbial kinetics without having an interest in the gradients, the homogeneous model would be sufficient for the investigated cultivation. However, the interest of this study was in the assessment and quantification of gradients.

In addition to the pH gradients (see earlier), the gradients of the main biological state variables were evaluated at 4 h of cultivation when the maximum growth rate was achieved and hence the gradients were most pronounced (Fig. 4.12). The distribution of the biomass, total lactic acid, and lactose concentrations predicted by the Monte Carlo simulation are shown in the seven compartments. The predictions in the compartments overlapped (Fig. 4.12 top row). Zooming to the mean concentrations in each compartment revealed that there existed small deviations of the biological state variable between the compartments (Fig. 4.12 bottom row). However, they were statistically not significant as the standard deviations of the state variables were larger than the deviations, namely $\sigma_{biomass} = 0.22$ g L^{-1}, $\sigma_{lactic\ acid} = 1.15$ g L^{-1}, and $\sigma_{lactose} = 4.15$ g L^{-1}. The small deviations were comprehensible: The biomass

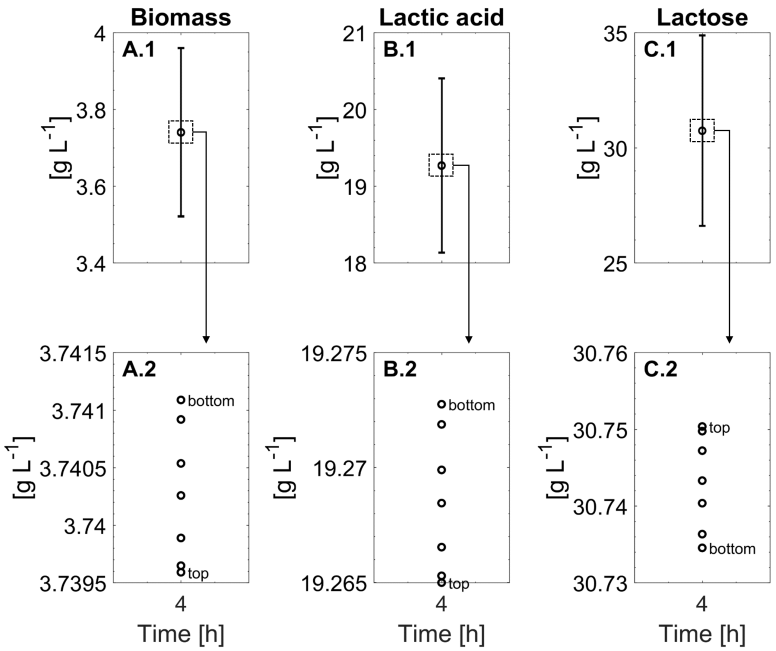


Figure 4.12: Gradients of the biological state variables. Mean concentrations (circles) with the standard deviation in the seven compartments of biomass (A), total lactic acid (B), and lactose (C). Reprinted from **Paper F.**

and lactic acid concentrations were higher in the bottom of the bioreactor as the pH was closer to pH_{opt} there, and the growth rate and lactic acid production rate were consequently higher. Accordingly, the lactose concentration was slightly lower in the bottom of the bioreactor.

The compartment model had a similar accuracy to the CFD model. The advantage of the compartment model was, however, that it could be solved in less than 2 s for one model input set. It is therefore desirable for on-line monitoring and control applications where the computational speed is required. Under the investigated conditions, there were pH gradients observed using both models but they had no influence on microbial growth. Gradients of the biological state variables were practically non-existent. However, if the cultivation conditions changed, e.g., a fed-batch cultivation was applied, these observations might change. It is therefore of interest to investigate fed-batch and continuous cultivations with the compartment model in terms of future work of this project.

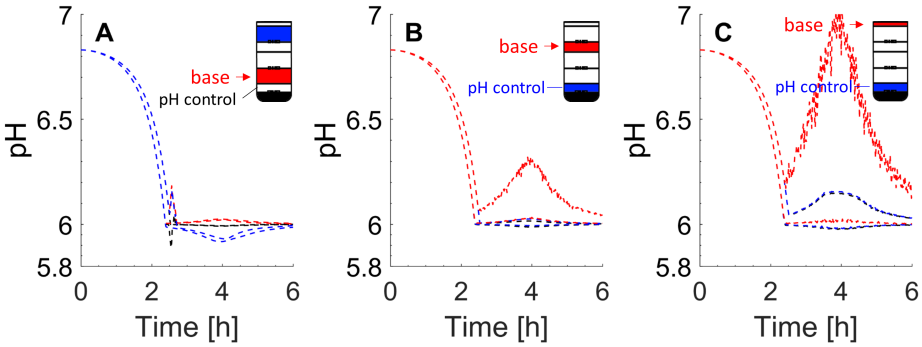


Figure 4.13: Testing of different base-addition locations. A: Ammonia solution addition below the middle impeller. B: Base addition below the top impeller. C: Base addition from the top of the bioreactor. Reprinted from **Paper F.**

Testing Different Base Addition Positions

As many lactic acid bacteria cultures are produced in batch mode, it was of interest in this study to demonstrate the compartment model simulation as a tool to assess different base-addition positions. It was desired to reduce the pH gradient in the cultivation even though the pH seemed not to have an effect on the investigated strain but might be critical for other lactic acid bacteria. Owing to the short simulation time in comparison to a CFD simulation, the compartment model was the preferred model to assess different scenarios. In this work, the location of the base addition was changed to three potential positions, namely under the middle impeller, under the top impeller, and the liquid surface at the top of the bioreactor. The pH sensor that was used as an input to the pH controller was kept at the same position as in the real setup in compartment 5 (the second compartment from the bottom).

In the simulation with the base addition below the middle impeller, a small pH gradient was predicted during the exponential growth phase (Fig. 4.13 A). The pH was 5.9 at the top of the bioreactor and 6.05 in the middle of the bioreactor, where the base was added. In the case the base addition was placed under the top impeller, the lowest pH would be 5.95 in the bottom and 6.3 in compartment 2, where the base was added (Fig. 4.13 B). Adding the base at the top of the bioreactor, as simulated with the addition of ammonia solution to compartment 7, a severe pH gradient was predicted (Fig. 4.13 C).

At the bottom of the bioreactor, the pH would not drop significantly under pH 6 as the pH sensor that is employed to control the pH is located closely. However, the pH at the top of the bioreactor was predicted to increase up to pH 7.

In order to decrease the pH gradients in the studied bioreactor, it is advised to place the base addition under the middle impeller. However, in general, when using a compartment model for process design or off-line optimization tasks, it is likely the best candidate should be simulated with a CFD simulation in the end to double-check the results and obtain comprehensive predictions.

If the pH would be controlled by adding base to the top of the bioreactor, the bacteria would have to adapt to changing pH conditions in the range of 1 pH unit while they circulate through the bioreactor because the extracellular pH affects the intracellular pH of the lactic acid bacteria, and hence their activity (Hansen et al., 2016). They maintain their intracellular pH typically with energy consuming antiporters (Sawatari and Yokota, 2007). As a result, the energy required for maintenance might increase, which could result in a lower biomass yield. These drastic pH changes could also result in a loss of cell viability. The investigation of the effects of oscillating extracellular pH on lactic acid bacteria should be addressed in future studies.

Taking the recent CFD studies in the literature and in this work together, CFD and compartment models are robust modelling tools for obtaining a qualitative and frequently even quantitative view of the cultivation system. In this work, an non-aerated bioreactor was investigated, and hence a one-phase simulation was performed. Future investigations should seek to model production-scale cultivations, including aerated fed-batch processes, and thus combine the presence of substrate, dissolved oxygen, and pH gradients.

The CFD and compartment models were successfully applied to consider fluid dynamics and predict the pH gradients in the *S. thermophilus* cultivation. For on-line applications and fast off-line tests, the compartment model is preferred because of the computational speed. Here, the compartment model was applied to test different base addition scenarios, and it will be applied for on-line monitoring and risk quantification in **Chapter 5**.

Chapter 5

On-line Risk Quantification

In Chapter 3, the probabilistic soft sensor has been described and validated for lab-scale cultivations. Chapter 4 has focused on the modelling of pH gradients. This final part of the work combines both tools, and demonstrates the monitoring of lactic acid bacteria in the 700-L bioreactor with a risk assessment tool. The CPP, pH gradient, and the CQA, biomass production, are evaluated.

This chapter is based upon the following article:

Paper G: Spann, R., Gernaey, K. V., & Sin, G. (2018). On-line Process Risk Assessment of a 700 L Lactic Acid Bacteria Cultivation. Frontiers. **Submitted**

5.1 Monitoring of pH Gradients

The compartment model, which was validated in **Chapter 4**, was used in the soft sensor, which was introduced with a single-compartment model in **Chapter 3**, to monitor a 700-L cultivation. The goal was to assess a CPP and CQA of the lactic acid bacteria cultivation, namely the pH gradient and biomass production, respectively.

The soft sensor was applied to a historical data set of a 700-L cultivation. The on-line data, namely the pH and quantity of added ammonia, were used as they would be available on-line in order to update the parameters in five min intervals. As described in **Chapter 3**, a Monte Carlo simulation with

200 input samples was performed to account for uncertainties in the model parameters, initial process conditions, and the base addition.

Fig. 5.1 presents the results of the soft sensor after 1, 2, 3, 5, and 6 h (in rows). The ammonia addition (Fig. 5.1, left column) and pH (Fig. 5.1, middle column) that was measured in the second compartment from the bottom were employed as inputs to the soft sensor. The soft sensor predicted both the measured and unmeasured state variables, whereas the biomass and lactose concentrations are shown in the right column of Fig. 5.1. The cultivation commenced with a pH of 6.8 (5.1, t = 1 h). As lactic acid was produced, the pH dropped to the control set point at pH 6, when the ammonia addition began. There was no gradient expected or predicted until the base addition started. The pH was therefore equal at all measurement and simulation points. The model predictions in the bottom compartment and top compartment no. 6 are shown. t_{lag} was updated using the pH measurements, and the improvement of the biomass predictions can be observed when comparing the prediction after 1 h (5.1, 1st row), 2 h (2nd row), and 3 h ((5.1, 3rd row). Of course, the off-line measurements of the biomass and lactose concentrations were not utilized to update the soft sensor, and are only shown for comparison's sake.

The ammonia addition started after ca. 2.5 h when the pH set point was reached (5.1, t = 3 h). A pH gradient between pH 5.8 and 6.1 was observed as measured and predicted in the previous chapter. The accuracy of the pH gradient has also been discussed in **Chapter 4**. The soft sensor used the quantity of added ammonia to update μ_{max}, and achieved an accurate prediction of the state variable (time = 3 - 6 h). The RMSSE of the biomass concentration was smaller than 0.4 $\pm$ 0.1 g L^{-1} corresponding to an error of less than 10 %.

The added value of the compartment model in the soft sensor is that – if applied on-line – the gradients can be assessed on-line, and therefore the risk of faulty batches could be diminished. The soft sensor provides insight into the bioreactor that can hardly be measured. Many large-scale bioreactors simply do not have ports for measurement devices at different positions. In case the pH gradient passes a critical level, operators could take appropriate

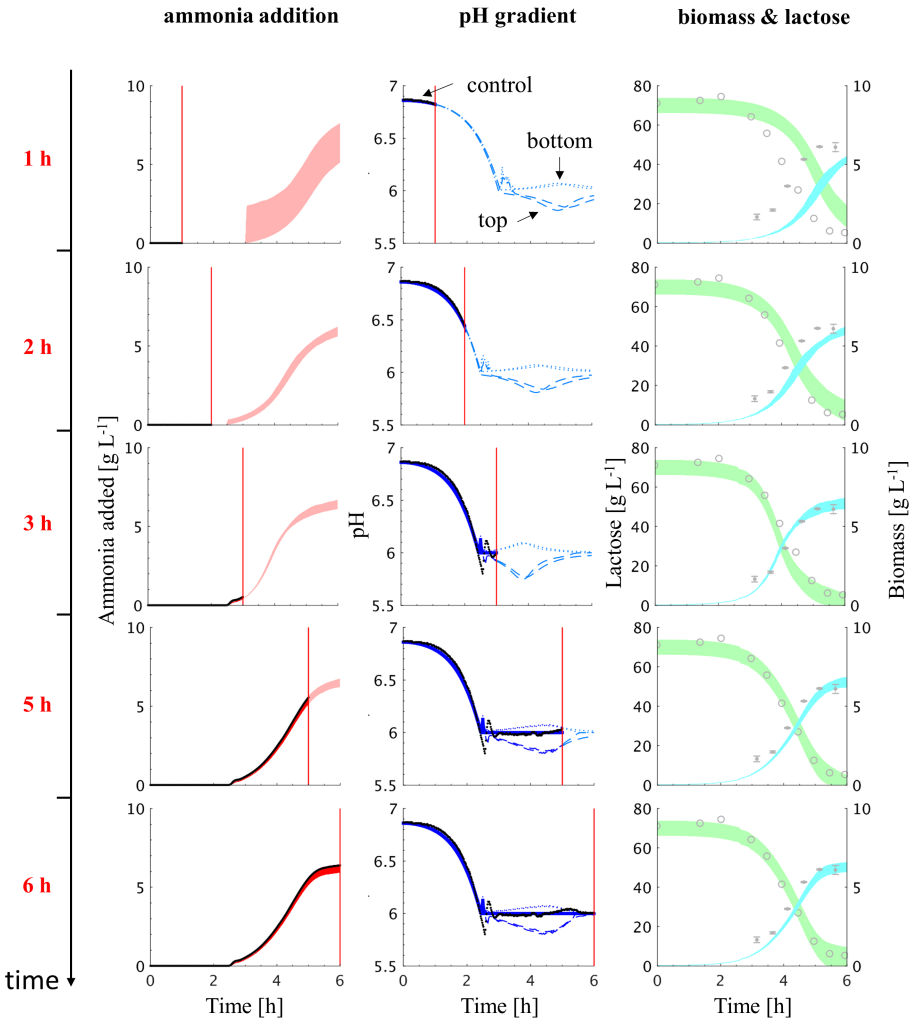


Figure 5.1: Soft sensor predictions of the 700-L *S. thermophilus* cultivation for risk-based decision making. The predictions at different time points (rows) are shown. The ammonia addition (left column), pH gradient (middle column), and biomass and lactose concentrations (right column) are shown. The on-line measurements (black dots) of the ammonia addition and pH were used as an input to the soft sensor. The off-line measurements of the biomass (grey dots with std. deviation) and the lactose (open circles) concentrations are shown for comparison only but not utilized to update the soft sensor. The 95 % confidence interval of the probabilistic model predictions are seen.

actions. There is also the possibility in the future to apply the soft sensor for automated on-line control or guidance of the operators (Jiménez-Hornero et al., 2009; Mears et al., 2017b). Actions to counteract a high pH gradient could include increasing the stirrer speed or changing the pH or temperature set point. A higher stirrer speed would improve the mixing and hence lead to a smaller pH gradient even despite restriction because shear stress must be considered (Arnaud et al., 1993; Lange et al., 2001). The set point of the pH or temperature could also be changed to a less optimal value in order to decrease cell growth. Less lactic acid would then be produced and the pH gradient might diminish. In the case severe pH gradients occur frequently, the bioreactor design should be improved, e.g., by altering the base inlet. As discussed in **Chapter 4**, the compartment and CFD model could support the design process.

5.2 Risk Quantification Using the Monte Carlo Simulation

The soft sensor provided a probabilistic distribution of the model outputs thanks to the Monte Carlo simulation. In this work, the goal was to produce a high biomass concentration as biomass is the product. From an industry firm's point of view, it would be of much added value to know the probability of not achieving the target biomass yield during cultivation. Furthermore, the risk could be quantified, i.e., how much biomass could be lost. The probability and risk would be indicators whether actions are needed to improve the process.

The risk of not producing the target biomass yield was calculated based on the results of the Monte Carlo simulation in the soft sensor. Here, a target yield of 0.09 ± 0.003 $g_{biomass}$ $g_{lactose}^{-1}$ was defined based on the yield that was achieved at the laboratory scale, and it was intended to obtain at least the same biomass yield when scaling up the process to 700 L. From the production perspective, the total biomass production per batch is also of interest as this is the amount that can be processed and sold. For the 700-L bioreactor, which was operated with 70 g L^{-1} lactose, the target production was defined as 4410 $g_{biomass}$ $batch^{-1}$ accordingly.

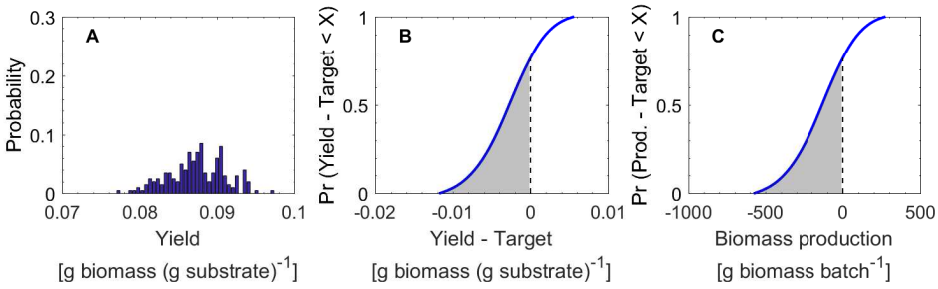


Figure 5.2: Probability distribution of the target biomass yield. Probability distribution of the predicted biomass yield obtained from the Monte Carlo simulation (A). Cumulative distribution function of the biomass yield (B). Cumulative distribution function of the total biomass production per batch (C). Reprinted from **Paper G**.

Methodology

The risk was quantified based on the likelihood that was obtained from the output of the Monte Carlo simulation. In this case, 200 model predictions were available.

Firstly, the probability distribution of the predicted biomass yield ($Y_{X/S}$) was calculated based on the predicted biomass concentration and initial lactose concentration:

$$Y_{X/S} = \frac{C_{X,predicted}}{C_{S,initial}} \tag{5.1}$$

The histogram in Fig. 5.2 A shows the probability distribution of the biomass yield based on the output of a Monte Carlo simulation as an example. The yield ranged from 0.076 to 0.096 $g_{biomass}\ g_{lactose}^{-1}$ in this example. Only a few predictions reached the target yield of 0.09 $g_{biomass}\ g_{lactose}^{-1}$.

Secondly, the cumulative distribution function was calculated (Fig. 5.2 B). The predicted yield was subtracted by the target yield in this step to easily differentiate between events that met the target and those that did not. The probability of not achieving the target yield was ca. 75 %. To put it another way, 75 % of the 200 model predictions did not reach the desired target biomass yield while 25 % were equal or larger than the target yield, accordingly.

In the third step, the yield was converted to the total biomass production per batch (Fig. 5.2 C). In the worst case, 500 $g_{biomass}$ $batch^{-1}$ could be lost in this cultivation, according to the simulation. However, the probability of this event was very low.

Finally, the risk was quantified. The risk of an undesired event (u.e.) is generally defined as the likelihood of the undesirable event multiplied with the consequence of this event (Cameron and Raman, 2005). The simulation that did not meet the target yield was defined as an undesired event in this work. To quantify the risk of all undesired events, the sum of the individual risks was calculated:

$$process\ risk = \sum_{m} consequence(u.e.) \cdot likelihood(u.e.) \tag{5.2}$$

The consequence was defined as the loss/plus of each of the j predictions (Eq. (5.3)). The likelihood of each event is obtained from the Monte Carlo simulation.

$$consequence_j = \hat{Y}_j - Y_{target} \tag{5.3}$$

where $\hat{Y}_j$ represents the yield predicted by the j^{th} simulation.

The risk is equivalent to the grey shaded area under to cumulative distribution function (Fig. 5.2 C). In the presented example, the risk was -140 $g_{biomass}$ $batch^{-1}$.

Risk Quantification in the Soft Sensor

Risk quantification was included in the monitoring system of the 700-L cultivation. The probability of not achieving the target biomass yield and the risk could therefore be obtained on-line each time the system was updated. In the current work, the biomass concentration that was predicted for the end of the cultivation after 6 h was used for the risk quantification. However, the methodology could be flexibly adjusted to other time points or other state variables.

Fig. 5.3 shows the risk that was predicted during the cultivation. In the beginning, the risk was high as much biomass could have been lost according to the model predictions. However, during the early stage of the cultivation, only a few on-line measurements were available to update the model parameters, and especially the pH was crucial to updating t_{lag} for an accurate model prediction as was found in **Chapter 3** (Fig. 3.5). It was therefore proposed to define an initialization phase for the risk quantification until t_{lag} had been finally updated (here: 2 h 25 min). This initialization phase should be adapted dynamically for each cultivation.

Once there was enough information for the soft sensor to provide an accurate prediction, the risk quantification was deemed to be reliable for the risk prediction phase. During the risk prediction phase, the risk oscillated between 0 and -140 $g_{biomass}$ $batch^{-1}$ in the investigated cultivation. The risk was close to zero at 2.5 h and 4.5 h, while at 3 h 45 min and at the end of the cultivation the risk was negative. The oscillating risk prediction might have been caused by the update procedure of μ_{max}. μ_{max} was updated with

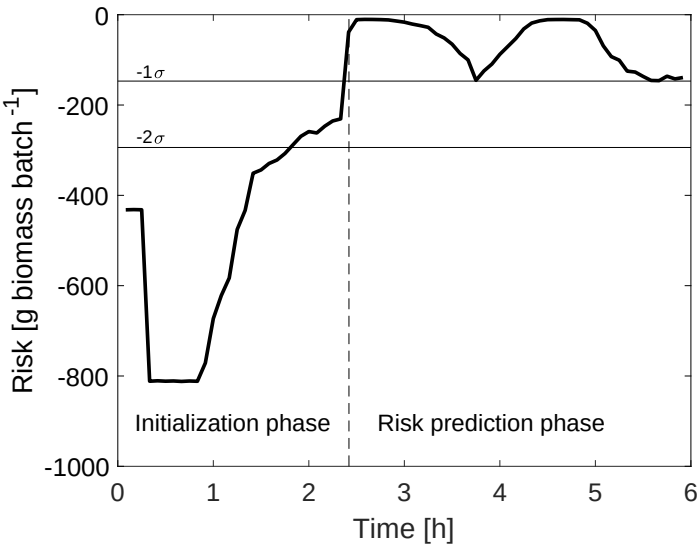


Figure 5.3: Risk quantification. The risk of how much biomass could be lost per batch was obtained from the soft sensor predictions. Limited on-line measurements were available for the soft sensor in the beginning of the cultivation (Initialization phase). With more on-line data, the risk could be predicted (Risk prediction phase). Reprinted from **Paper G**.

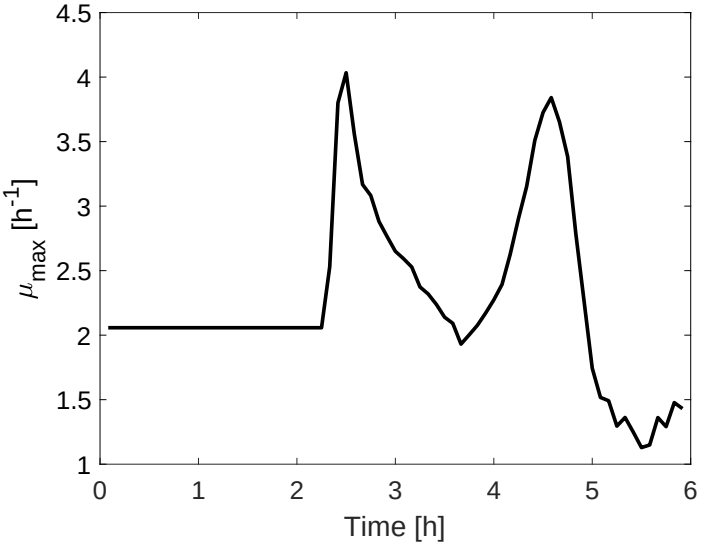


Figure 5.4: μ_{max} value updated in the soft sensor. Reprinted from **Paper G**.

the quantity of added ammonia in the data reconciliation and parameter update step (**Chapter 3**). This led to dynamic changes in μ_{max} (Fig. 5.4), and hence when μ_{max} was low, the risk of losing biomass was higher. Nevertheless, the predicted risk was within the 68 % confidence interval (1σ) that was estimated from the lab-scale data. The risk was therefore seen as natural variability, and there was no yield reduction observed for this cultivation compared to the lab-scale experiments. Nevertheless, additional lab- and pilot-scale experiments would be needed to validate the yield statistically.

This on-line risk quantification could expand the traditional risk assessment that is conducted for biotechnological processes. In addition to a risk assessment that is conducted before the process is run, this method could provide a risk measure for the product loss in real time. Further uncertainties could be included, and hence a higher degree of reliability of the risk quantification would be achieved. The methodology could furthermore be extended to calculate the monetary value of the product in order to consider the economic side, as well, as the production is profit-oriented at the end of the day.

The soft sensor was successfully applied as a PAT risk assessment tool. It was demonstrated with a pilot-scale cultivation, where heterogeneities exist.

Specifically, the key CPP, namely the pH gradient, could be predicted. Having such information on-line would enable plant operators to evaluate the mixing strategy and ammonia addition, and react accordingly. Furthermore, the CQA, the final biomass production, was assessed based on the probabilistic predictions of the Monte Carlo simulation. The operation strategy and decision to alter process parameters could therefore be based on the predicted risk.

Conclusions and Future Perspectives

In the work presented herein, it was demonstrated that pH gradients and process risks can be predicted and assessed during lactic acid bacteria cultivations. The mechanistic modelling framework was applied to a *S. thermophilus* cultivation and it was shown that the applied tools could support multiple phases of bioprocess development. Specifically, the framework was applied as a PAT tool for risk-based on-line monitoring, it improved process understanding, and it proposed a better process design. To accomplish this, the biokinetic model was coupled with a chemical model for pH calculation, and then combined with CFD and compartment models. All modelling modules were experimentally validated. The combination of the model with fluid dynamics, and the application as a probabilistic soft sensor contributed to risk-based process analytical technologies, which are highly desired for bioprocesses nowadays. The model-based soft sensor provided a PAT tool to monitor the progress of the cultivation, especially the pH as a key parameter in lactic acid bacteria batch cultivations. This enables achievement of virtual insight into bioreactors, almost similar to creating an opportunity by looking through a window into them.

The soft sensor supplied insight into the biological variables, which are more comprehensible than physical parameters, such as pump speed, that are traditionally monitored. The soft sensor improved the prediction accuracy of the biomass growth because it updated the lag-time parameter. It required merely the already installed measurement equipment of pH and base addition quantity in contrast to other soft sensors that are, for example, based on spectroscopic measurements.

This PAT tool could also be used for scheduling of the different unit operations ensuring a lean, on-time production as it is capable of predicting the time when the target yield will be reached. The other unit operations, in particular the downstream units, could be scheduled and prepared in time, thereby ensuring their efficient application. In addition, other tasks, such as cleaning in place, sterilization in place, media preparation, and preparation of pre-cultures could be optimized when knowing the completion time of the cultivation. An optimized schedule with a reduced downtime of production units will increase productivity and profit.

Robust model predictions were achieved by the application of a Monte Carlo simulation considering input uncertainties in the model parameters, the initial process conditions, and measurements. Firstly, the comprehensive assessment of the model parameter estimation accuracy was conducted in order to determine the limitations of the model, which are inevitable owing to the quality of the data that was used for parameter estimation. In addition to the uncertainties in the model parameters, variations of the carbon source and the biomass inoculum concentration, as well as measurement inaccuracies, were considered in the Monte Carlo simulation because they vary naturally from batch to batch. In the future, further stochastic, structural, and input uncertainties should be considered. Uncertainties in the chemical model could be considered, for instance, together with random failure of equipment and uncertainties in the model structure.

The probabilistic model predictions allowed for a risk-based monitoring and more objective interpretation of the model outputs compared to deterministic model simulations. The risk of not achieving target productivity was quantified and could be updated on-line within the soft sensor. A natural progression of this work is to incorporate economic aspects in risk calculation as the production is ultimately economically driven. In this way, the process operators would be equipped with a tool that allows risk-based decision making on the basis of profit/loss in monetary terms. Furthermore, the soft sensor could detect process failures early and alert process operators immediately. If this soft-sensor would be integrated at the production site, the operators could be alerted to process deviations in a timely fashion. If, for example, biomass growth stops, the soft sensor together with physical measurements

like pressure, temperature, and flow rate measurements could support the operators to find the cause of failure.

The soft sensor could also be used for automated process control in the future. An automated control algorithm could adjust process parameters in order to optimize the process, i.e., minimize the risk of faulty batches, minimize loss of product, maximize the profit, and optimize the quality and quantity of the product. In the more distant future, this tool could then be included in a plant-wide on-line monitoring, control, and scheduling platform that handles all up- and downstream units in an integrated manner.

With respect to modelling pH gradients in lactic acid bacteria cultivations, two approaches have been considered: CFD modelling and compartment modelling. Both predicted the spatial pH gradients as they were measured with multi-position pH sensors. The predictions allowed judging the importance of certain process parameters, such as the stirrer speed or the dosing position, in the production process. However, the CFD and compartment models had limitations. Firstly, the high computational demand of the CFD simulations limited their ability to be implemented in on-line process monitoring and control. Secondly, the data-driven linear pH correlation limited the predictability of pH in the CFD simulation to the validated range of the pH correlation. The compartment model overcame those limitations but decreased spatial resolution and accuracy. Local heterogeneities – as, for example, the high pH close to the base inlet observed in this study – are not visible with the compartment model. The choice between the CFD and compartment model depends therefore on the application. Ideally, the aim would be for the on-line monitoring to implement the soft sensor with the compartment model at the production site as it provides the required speed combined with sufficient accuracy in the present case. The pH could then be assessed on-line and reduced if it passed beyond a critical value, for example, by increasing the stirrer speed.

There is huge potential for scale-down experiments that are designed based on CFD simulations. Future studies could therefore benefit from the results, e.g., during strain screening, medium testing, and process design, and use a bioreactor setup at the laboratory scale that mimics heterogeneous

production-scale conditions. On the one hand, multi-compartment experiments, e.g., consisting of stirred-tank bioreactors and plug-flow reactors could be used. On the other hand, high-throughput experiments employing liquid handling stations could be designed. In the case of the presented *S. thermophilus* cultivation, pH gradients could, for example, be mimicked in a bench-scale three-compartment setup, where the largest bioreactor is controlled to a pH 6, ammonia is added to the second compartment to mimic the bottom base-addition zone of the bioreactor, and a third reactor that resembles the top zone of the bioreactor, where the pH drops owing to lactic acid secretion. However, as pumps are used to pump the liquid continuously back and forth between the bioreactors, experiments are needed to ensure that pumping has no influence on the results as one is mainly interested in the effect of the gradients. Microorganisms could experience high shear rates in the pump, which could affect microbial response, and gases like CO_2 could be introduced in the liquid phase that distort the results, among other influential factors. In a liquid-handling station using smaller volumes, pulse-feeding experiments with acids and bases could be performed to resemble heterogeneous large-scale conditions.

The process design phase should, in general, be accompanied by simulations that consider fluid dynamics and the resulting gradients. Simulations allow testing different scenarios. In the case of lactic acid bacteria, it might be of interest to test varying stirrer speeds and assess the resulting pH gradients. This is especially functional at a large scale seeing that multi-position measurements are difficult to perform during production owing to experimental challenges and regulatory requirements. In this work, a better dosing position for the base was predicted to decrease the pH gradient. Nevertheless, an experimental validation is still necessary to be performed.

The developed mechanistic model was capable of predicting a continuous process even though it was calibrated with batch data. With the biotechnology industry generally interested in moving to continuous production processes, the value of a mechanistic model in terms of extrapolation capabilities was shown. The development time that is needed for a mechanistic model and is regarded as a drawback of these models might pay off when the model aids the design of new processes – the number of costly experiments that are

required to create an entirely new process could be reduced.

The ultimate goal here, which is to understand and predict the cultivation process accurately, requires further research. In a large-scale process, population heterogeneities emerge as each cell has its own history, e.g., its age and trajectory trough the bioreactor. The utilization of single-cell measurements, such as on-line flow cytometry, would provide additional information about the process as, for example, the viability of individual cells could be assessed. Several studies are currently investigating on important research areas needed in an integrated approach for successful bioprocess development. To name a few: i) the impact of gradients on different strains; ii) population heterogeneities; iii) agent-based modelling; iv) CFD; and v) virtual reality. Finally, simulations that model the different -omics levels and track the lifeline of individual cells in combination with fluid dynamic models that capture gradients are required. Such simulations could serve as a virtual window into the bioreactor, displaying unmeasured parameters and support the different phases of bioprocess development.

Abbreviations and Nomenclature

Abbreviations

Abbreviation	Definition
CFD	Computational fluid dynamics
CHO	Chinese hamster ovary
CM	Compartment model
CPP	Critical process parameter
CQA	Critical quality attribute
DTU	Technical University of Denmark
GMP	Good manufacturing practice
HPLC	High-performance liquid chromatography
LAB	Lactic acid bacteria
LHS	Latin Hypercube Sampling
ODE	Ordinary differential equation
PAT	Process analytical technology
PDE	Partial differential equation
RMSSE	Root mean sum of squared error

Symbols

Symbol	Definition
C_{H^+}	H^+ concentration (mol L^{-1})
C_{LA}	lactate concentration (g L^{-1})
C_{OH^-}	OH^- concentration (mol L^{-1})
C_P	total lactic acid (lactate and lactic acid) concentration (g L^{-1})
C_S	lactose (substrate) concentration (g L^{-1})
C_X	biomass concentration (g L^{-1})
$Error_i$	weighted model prediction error at time point i
f_{lag}	lag-time function (-)
f_P	lactic acid inhibition function (-)
f_{pH}	pH dependency function (-)
f_S	substrate limitation and inhibition function (-)
$H_2CO_3^*$	dissolved CO_2 and H_2CO_3
I	ionic strength g L^{-1}
K'_{C1}	apparent equilibrium constant for the carbonic acid system (-)
K_I	substrate inhibition parameter (g L^{-1})
K_{La}	lactate inhibition parameter (g L^{-1})
K_{La1}	pH dependent lactate inhibition parameter (g L^{-1})
K'_{NH}	apparent equilibrium constant for the ammonia system (-)
K_P	P-controller controller gain
$K_{P,La}$	2. lactate inhibition parameter (L g^{-1})
$K_{P,pH1}$	lactate inhibition pH parameter (-)
$K_{P,pH2}$	2. lactate inhibition pH parameter (-)
K'_{P1}	apparent equilibrium constant for the H_3PO_4 system (-)
K'_{P2}	apparent equilibrium constant for the $H_2PO_4^-$ system (-)
$K'_{r,C1}$	apparent reverse rate constant for H_2CO_3 dissociation (s^{-1})
$K'_{r,LA}$	apparent reverse rate constant for lactic acid dissociation (s^{-1})

continued on the next page

continued from the previous page

Symbol	Definition
$K'_{r,NH}$	apparent reverse rate constant for NH_4 dissociation (s^{-1})
$K'_{r,P1}$	apparent reverse rate constant for H_3PO_4 dissociation (s^{-1})
$K'_{r,P2}$	apparent reverse rate constant for $H_2PO_4^-$ dissociation (s^{-1})
$K'_{r,W}$	apparent reverse rate constant for water dissociation (s^{-1})
K'_Z	apparent equilibrium constant for the unspecified compound system (-)
$log\ D^2$	logarithmic squared deviation of the normalized pH values
n	number of measurement points
pH'_i	normalized pH of the i^{th} sensor
pH_{opt}	optimal pH parameter in the pH function (-)
pH_{set}	pH control set point (-)
q_{gal}	volumetric galactose secretion rate (C-mol L^{-1} h^{-1})
q_{NH}	volumetric ammonia consumption rate (mol L^{-1} h^{-1})
$q_{NH,add}$	volumetric ammonia addition rate (mol L^{-1} h^{-1})
q_P	volumetric lactic acid secretion rate (C-mol L^{-1} h^{-1})
q_{Ph}	volumetric phosphoric acid consumption rate (mol L^{-1} h^{-1})
q_S	volumetric substrate consumption rate (C-mol L^{-1} h^{-1})
q_X	volumetric biomass growth rate (C-mol L^{-1} h^{-1})
RE	relative error (-)
RMSSE	root mean sum of squared error (g L^{-1})
T	temperature in the cultivation broth (K)
t	time variable (h)
u.e.	undesired event
$\hat{Y}$	predicted yield
Y_{gal}	galactose yield (g g^{-1})
$Y_{X/S}$	biomass yield (g g^{-1})
z_i	charge number of the i^{th} ion
$\hat{y}_i$	model output of the i^{th} state variable

Greek symbols

Symbol	Definition
α	growth related production coefficient of lactic acid (g g^{-1})
$\hat{\theta}_i$	estimated value of the i^{th} parameter
μ_{max}	maximum specific growth rate (h^{-1})
σ	standard deviation
σ_{pH}	spread parameter in the gaussian pH function (-)
$\sigma_{\hat{\theta}_i}$	standard deviation of the i^{th} estimated parameter

Bibliography

Acuña, G., E. Latrille, C. Béal, G. Corrieu, and A. Chéruy
1994. On-line estimation of biological variables during pH controlled lactic acid fermentations. *Biotechnology and Bioengineering*, 44(10):1168–1176.

Adam, S., D. Suzzi, C. Radeke, and J. G. Khinast
2011. An integrated Quality by Design (QbD) approach towards design space definition of a blending unit operation by Discrete Element Method (DEM) simulation. *European Journal of Pharmaceutical Sciences*, 42(1-2):106–115.

Aghababaie, M., M. Khanahmadi, and M. Beheshti
2015. Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures. *Food and Bioproducts Processing*, 94(April):657–667.

Åkerberg, C., K. Hofvendahl, B. Hahn-Hägerdal, and G. Zacchi
1998. Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour. *Applied Microbiology and Biotechnology*, 49(6):682–690.

Albaek, M. O., K. V. Gernaey, M. S. Hansen, and S. M. Stocks
2011. Modeling enzyme production with Aspergillus oryzae in pilot scale vessels with different agitation, aeration, and agitator types. *Biotechnology and Bioengineering*, 108(8):1828–1840.

Almquist, J., M. Cvijovic, V. Hatzimanikatis, J. Nielsen, and M. Jirstrand
2014. Kinetic models in industrial biotechnology – Improving cell factory performance. *Metabolic Engineering*, 24:38–60.

Altiok, D., F. Tokatlı, and e. Harsa
2006. Kinetic modelling of lactic acid production from whey byLactobacillus casei (NRRL B-441). *Journal of Chemical Technology & Biotechnology*, 81(7):1190–1197.

Alvarado, A., S. Vedantam, P. Goethals, and I. Nopens
2012. A compartmental model to describe hydraulics in a full-scale waste stabilization pond. *Water Research*, 46(2):521–530.

Alves-Rausch, J., R. Bienert, C. Grimm, and D. Bergmaier
2014. Real time in-line monitoring of large scale Bacillus fermentations with near-infrared spectroscopy. *Journal of Biotechnology*, 189:120–128.

Amanullah, A., C. M. McFarlane, A. N. Emery, and A. W. Nienow
2001. Scale-down model to simulate spatial pH variations in large-scale bioreactors. *Biotechnology and Bioengineering*, 73(5):390–399.

Amrane, A. and Y. Prigent
1998. Influence of an initial addition of lactic acid on growth, acid production and their coupling for batch cultures of Lactobacillus helveticus. *Bioprocess Engineering*, 19(4):307–312.

Arnaud, J., C. Lacroix, C. Foussereau, and L. Choplin
1993. Shear stress effects on growth and activity of Lactobacillus delbrueckii subsp. bulgaricus. *Journal of Biotechnology*, 29(1-2):157–175.

Baez, A., N. Flores, F. Bolívar, and O. T. Ramírez
2011. Simulation of dissolved CO2 gradients in a scale-down system: A metabolic and transcriptional study of recombinant Escherichia coli. *Biotechnology Journal*, 6(8):959–967.

Baldinger, A., L. Clerdent, J. Rantanen, M. Yang, and H. Grohganz
2011. Quality by design approach in the optimization of the spray-drying process. *Pharmaceutical Development and Technology*, 17(4):389–397.

Bannari, R., A. Bannari, P. Vermette, and P. Proulx
2012. A model for cellulase production from Trichoderma reesei in an airlift reactor. *Biotechnology and Bioengineering*, 109(8):2025–2038.

Barua, S., X. Gao, H. Pasman, and M. S. Mannan
2016. Bayesian network based dynamic operational risk assessment. *Journal of Loss Prevention in the Process Industries*, 41:399–410.

Bashiri, H., F. Bertrand, and J. Chaouki
2016. Development of a multiscale model for the design and scale-up of gas/liquid stirred tank reactors. *Chemical Engineering Journal*, 297:277–294.

Belini, V. L., P. Wiedemann, and H. Suhr
2013. In situ microscopy: A perspective for industrial bioethanol production monitoring. *Journal of Microbiological Methods*, 93(3):224–232.

Bezzo, F. and S. Macchietto
2004. A general methodology for hybrid multizonal/CFD models: Part II. Automatic zoning. *Computers and Chemical Engineering*, 28(4):513–525.

Bezzo, F., S. Macchietto, and C. C. Pantelides
2003. General hybrid multizonal/CFD approach for bioreactor modeling. *AIChE Journal*, 49(8):2133–2148.

Biazar, J., M. Tango, E. Babolian, and R. Islam
2003. Solution of the kinetic modeling of lactic acid fermentation using Adomian decomposition method. *Applied Mathematics and Computation*, 144(2-3):433–439.

Biechele, P., C. Busse, D. Solle, T. Scheper, and K. Reardon
2015. Sensor systems for bioprocess monitoring. *Engineering in Life Sciences*, 15(5):469–488.

Blombach, B., J. Buchholz, T. Busche, J. Kalinowski, and R. Takors
2013. Impact of different CO2/HCO3- levels on metabolism and regulation in Corynebacterium glutamicum. *Journal of Biotechnology*, 168(4):331–340.

Bluma, A., T. Höpfner, P. Lindner, C. Rehbock, S. Beutel, D. Riechers, B. Hitzmann, and T. Scheper
2010. In-situ imaging sensors for bioprocess monitoring: state of the art. *Analytical and Bioanalytical Chemistry*, 398(6):2429–2438.

Boonmee, M., N. Leksawasdi, W. Bridge, and P. L. Rogers
2003. Batch and continuous culture of Lactococcus lactis NZ133: Experimental data and model development. *Biochemical Engineering Journal*, 14(2):127–135.

Bouguettoucha, A., B. Balannec, and A. Amrane
2011. Unstructured models for lactic acid fermentation - a review. *Food Technology and Biotechnology*, 49(1):3–12.

Buchholz, J., M. Graf, A. Freund, T. Busche, J. Kalinowski, B. Blombach, and R. Takors
2014. CO2/HCO3- perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in Corynebacterium glutamicum. *Applied Microbiology and Biotechnology*, 98(20):8563–8572.

Burgos-Rubio, C. N., M. R. Okos, and P. C. Wankat
2000. Kinetic study of the conversion of different substrates to lactic acid using Lactobacillus bulgaricus. *Biotechnology progress*, 16(3):305–314.

Bylund, F., E. Collet, S.-O. Enfors, and G. Larsson
1998. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Engineering*, 18(3):171.

Cachon, R. and C. Diviès
1993. Modeling of growth and lactate fermentation by Lactococcus lactis subsp. lactis biovar. diacetylactis in batch culture. *Applied Microbiology and Biotechnology*, 40(1):28–33.

Cameron, I. and R. Raman
2005. *Process Systems Risk Management*. Elsevier Academic Press.

Campbell, K., J. Xia, and J. Nielsen
2017. The Impact of Systems Biology on Bioprocessing. *Trends in Biotechnology*, 35(12):1156–1168.

Chis, O.-T., J. R. Banga, and E. Balsa-Canto
2011. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. *PLoS ONE*, 6(11):e27755.

Cortés, J. T., N. Flores, F. Bolívar, A. R. Lara, and O. T. Ramírez
2016. Physiological effects of pH gradients on Escherichia coli during plasmid DNA production. *Biotechnology and Bioengineering*, 113(3):598–611.

Craven, S., J. Whelan, and B. Glennon
2014. Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *Journal of Process Control*, 24(4):344–357.

Davies, C. W.
1962. *Ion Association*. Londond: Butterworth.

de Jonge, L. P., N. A. Buijs, A. ten Pierick, A. Deshmukh, Z. Zhao, J. A. Kiel, J. J. Heijnen, and W. M. van Gulik
2011. Scale-down of penicillin production in Penicillium chrysogenum. *Biotechnology Journal*, 6(8):944–958.

Delafosse, A., S. Calvo, M.-L. Collignon, F. Delvigne, M. Crine, and D. Toye
2015. Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – Comparison to experimental flow characterization and particle tracking. *Chemical Engineering Science*, 134:457–466.

Delafosse, A., M.-L. Collignon, S. Calvo, F. Delvigne, M. Crine, P. Thonart, and D. Toye
2014. CFD-based compartment model for description of mixing in bioreactors. *Chemical Engineering Science*, 106:76–85.

Delafosse, A., F. Delvigne, M. L. Collignon, M. Crine, P. Thonart, and D. Toye
2010. Development of a compartment model based on CFD simulations for description of mixing in bioreactors. *Biotechnology, Agronomy and Society and Environment*, 14(SPEC. ISSUE 2):517–522.

Delvigne, F., J. Destain, and P. Thonart
2006. A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. *Biochemical Engineering Journal*, 28(3):256–268.

Delvigne, F., R. Takors, R. Mudde, W. van Gulik, and H. Noorman
2017. Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. *Microbial Biotechnology*, 10(5):1267–1274.

Djuris, J. and Z. Djuric
2017. Modeling in the quality by design environment: Regulatory requirements and recommendations for design space and control strategy appointment. *International Journal of Pharmaceutics*, 533(2):346–356.

Elqotbi, M., S. D. Vlaev, L. Montastruc, and I. Nikov
2013. CFD modelling of two-phase stirred bioreaction systems by segregated solution of the Euler–Euler model. *Computers & Chemical Engineering*, 48:113–120.

Enfors, S. O., M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jürgen, E. Krüger, T. Schweder, G. Hamer, D. O'Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. McFarlane, A. Nienow, T. Kovacs, C. Trägårdh, L. Fuchs, J. Revstedt, P. C. Friberg, B. Hjertager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H. Y. Lin, P. Neubauer, R. Van der Lans, K. Luyben, P. Vrabel, and Å. Manelius
2001. Physiological responses to mixing in large scale bioreactors. *Journal of Biotechnology*, 85(2):175–185.

Esener, a. a., J. a. Roels, and N. W. Kossen
1981. The Influence of Temperature on the Maximum Specific Growth Rate of Klebsiella pneumoniae. *Biotechnology and bioengineering*, XXIII(3):1401–1405.

Esener, A. A., J. A. Roels, and N. W. F. Kossen
1983. Theory and applications of unstructured growth models: Kinetic and energetic aspects. *Biotechnology and Bioengineering*, 25(12):2803–2841.

Fayolle, P., D. Picque, and G. Corrieu
1997. Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy. *Vibrational Spectroscopy*, 14(2):247–252.

Fernandes, R. L., V. K. Bodla, M. Carlquist, A.-L. Heins, A. E. Lantz, G. Sin, and K. V. Gernaey
2012. Applying Mechanistic Models in Bioprocess Development. In *Advances in Biochemical Engineering/Biotechnology*, Pp. 137–166. Springer Berlin Heidelberg.

Ferrari, A., S. Gutiérrez, and G. Sin
2016. Modeling a production scale milk drying process: parameter estimation, uncertainty and sensitivity analysis. *Chemical Engineering Science*, 152:301–310.

Flores-Alsina, X., K. Solon, C. K. Mbamba, S. Tait, K. V. Gernaey, U. Jeppsson, and D. J. Batstone
2016. Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes. *Water Research*, 95:370–382.

Frahm, B.
2013. Seed Train Optimization for Cell Culture. In *Animal Cell Biotechnology*, Pp. 355–367. Humana Press.

Fu, W. and a. P. Mathews
1999. Lactic acid production from lactose by Lactobacillus plantarum: Kinetic model and effects of pH, substrate, and oxygen. *Biochemical Engineering Journal*, 3(3):163–170.

García-Muñoz, S., C. V. Luciani, S. Vaidyaraman, and K. D. Seibert
2015. Definition of Design Spaces Using Mechanistic Models and Geometric Projections of Probability Maps. *Organic Process Research and Development*, 19(8):1012–1023.

George, S., G. Larsson, and S. O. Enfors
1993. A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of Saccharomyces cerevisiae. *Bioprocess Engineering*, 9(6):249–257.

George, S., G. Larsson, K. Olsson, and S.-O. Enfors
1998. Comparison of the Baker's yeast process performance in laboratory and production scale. *Bioprocess Engineering*, 18(2):135.

Gernaey, K. V., A. E. Cervera-Padrell, and J. M. Woodley
2012. A perspective on PSE in pharmaceutical process development and innovation. *Computers & Chemical Engineering*, 42:15–29.

Gernaey, K. V., A. E. Lantz, P. Tufvesson, J. M. Woodley, and G. Sin
2010. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in Biotechnology*, 28(7):346–354.

Golabgir, A. and C. Herwig
2016. Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production. *Chemie Ingenieur Technik*, 88(6):764–776.

Goldrick, S., A. Ştefan, D. Lovett, G. Montague, and B. Lennox
2015. The development of an industrial-scale fed-batch fermentation simulation. *Journal of Biotechnology*, 193:70–82.

Guha, D., M. P. Dudukovic, P. A. Ramachandran, S. Mehta, and J. Alvare
2006. CFD-based compartmental modeling of single phase stirred-tank reactors. *AIChE Journal*, 52(5):1836–1846.

Halton, J. H.
1964. Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702.

Hansen, G., C. L. Johansen, G. Marten, J. Wilmes, L. Jespersen, and N. Arneborg
2016. Influence of extracellular pH on growth, viability, cell size, acidification activity, and intracellular pH of Lactococcus lactis in batch fermentations. *Applied Microbiology and Biotechnology*, 100(13):5965–5976.

Haringa, C., A. T. Deshmukh, R. F. Mudde, and H. J. Noorman
2017. Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chemical Engineering Science*, 170:653–669.

Haringa, C., W. Tang, A. T. Deshmukh, J. Xia, M. Reuss, J. J. Heijnen, R. F. Mudde, and H. J. Noorman
2016. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Engineering in Life Sciences*, 16(7):652–663.

Helton, J. C. and F. J. Davis
2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69.

Henze, M., W. Gujer, T. Mino, and M. Lossdreacht
2000. *Activated sludge models ASM1, ASM2, ASM2d and ASM3*. IWA Publishing Company.

ICH Q9
2005. Quality Risk Management.

Iman, R. L. and W. J. Conover
1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311–334.

Jiménez-Hornero, J. E., I. M. Santos-Dueñas, and I. García-García
2009. Optimization of biotechnological processes. The acetic acid fermentation. Part III: Dynamic optimization. *Biochemical Engineering Journal*, 45(1):22–29.

Johnsson, O.
2015. *Perturbation-based Control of Industrial Fed-batch Bioprocesses*. Phd thesis, Lund University.

Kadlec, P., B. Gabrys, and S. Strandt
2009. Data-driven Soft Sensors in the process industry. *Computers & Chemical Engineering*, 33(4):795–814.

Kar, T., F. Delvigne, M. Masson, J. Destain, and P. Thonart
2008. Investigation of the effect of different extracellular factors on the

lipase production by Yarrowia lipolityca on the basis of a scale-down approach. *Journal of Industrial Microbiology & Biotechnology*, 35(9):1053–1059.

Karpinska, A. M. and J. Bridgeman
2016. CFD-aided modelling of activated sludge systems – A critical review. *Water Research*, 88:861–879.

Käß, F., S. Junne, P. Neubauer, W. Wiechert, and M. Oldiges
2014. Process inhomogeneity leads to rapid side product turnover in cultivation of Corynebacterium glutamicum. *Microbial Cell Factories*, 13(1):6.

Konakovsky, V., C. Clemens, M. M. Müller, J. Bechmann, and C. Herwig
2017. A robust feeding strategy to maintain set-point glucose in mammalian fed-batch cultures when input parameters have a large error. *Biotechnology Progress*, 33(2):317–336.

Kroll, P., A. Hofer, I. V. Stelzer, and C. Herwig
2017. Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. *Process Biochemistry*, 62:24–36.

Kuschel, M., F. Siebler, and R. Takors
2017. Lagrangian Trajectories to Predict the Formation of Population Heterogeneity in Large-Scale Bioreactors. *Bioengineering*, 4(2):27.

Lange, H., P. Taillandier, and J.-P. Riba
2001. Effect of high shear stress on microbial viability. *Journal of Chemical Technology & Biotechnology*, 76(5):501–505.

Langheinrich, C. and A. W. Nienow
1999. Control of pH in large-scale, free suspension animal cell bioreactors: Alkali addition and pH excursions. *Biotechnology and Bioengineering*, 66(3):171–179.

Lapin, A., D. Müller, and M. Reuss
2004. Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler-Lagrange Methods:∼ Traveling along the Lifelines of Single Cells†. *Industrial & Engineering Chemistry Research*, 43(16):4647–4656.

Lapin, A., J. Schmid, and M. Reuss
2006. Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor—A structured–segregated approach. *Chemical Engineering Science*, 61(14):4783–4797.

Lara, A. R., E. Galindo, O. T. Ramírez, and L. A. Palomares
2006. Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells. *Molecular Biotechnology*, 34(3):355–382.

Lemoine, A., M. H. Limberg, S. Kästner, M. Oldiges, P. Neubauer, and S. Junne
2016. Performance loss of Corynebacterium glutamicum cultivations under scale-down conditions using complex media. *Engineering in Life Sciences*, 16(7):620–632.

Li, P., H. Arellano-Garcia, and G. Wozny
2008. Chance constrained programming approach to process optimization under uncertainty. *Computers & Chemical Engineering*, 32(1-2):25–45.

Limberg, M. H., J. Schulte, T. Aryani, R. Mahr, M. Baumgart, M. Bott, W. Wiechert, and M. Oldiges
2017. Metabolic profile of 1,5-diaminopentane producing Corynebacterium glutamicum under scale-down conditions: Blueprint for robustness to bioreactor inhomogeneities. *Biotechnology and Bioengineering*, 114(3):560–575.

Loewenthal, R. E., G. A. Ekama, and G. R. Marais
1989. Mixed weak acid/base systems Part I - Mixture characterisation. *Water SA*, 15(1):3–24.

Löffler, M., J. D. Simen, J. Müller, G. Jäger, S. Laghrami, K. Schäferhoff, A. Freund, and R. Takors
2017. Switching between nitrogen and glucose limitation: Unraveling transcriptional dynamics in Escherichia coli. *Journal of Biotechnology*, 258:2–12.

Luedeking, R. and E. L. Piret
1959. Kinetic study of the lactic acid fermentation. Batch process at controlled pH. *Biotechnology and Bioengineering*, 67(6):636–644.

Luttmann, R., D. G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V. C. Hass, C. Kaiser, C. Preusse, G. Striedner, and C.-F. Mandenius
2012. Soft sensors in bioprocessing: A status report and recommendations. *Biotechnology Journal*, 7(8):1040–1048.

Mandenius, C.-F. and R. Gustavsson
2014. Mini-review: soft sensors as means for PAT in the manufacture of biotherapeutics. *Journal of Chemical Technology & Biotechnology*, 90(2):215–227.

Markopoulos, J., C. Christofi, and I. Katsinaris
2007. Mass Transfer Coefficients in Mechanically Agitated Gas-Liquid Contactors. *Chemical Engineering & Technology*, 30(7):829–834.

Marquard, D., C. Schneider-Barthold, S. Düsterloh, T. Scheper, and P. Lindner
2017. Online monitoring of cell concentration in high cell density Escherichia coli cultivations using in situ Microscopy. *Journal of Biotechnology*, 259:83–85.

McKay, M. D., R. J. Beckman, and W. J. Conover
1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245.

McKay, M. D., J. D. Morrison, and S. C. Upton
1999. Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, 117(1-2):44–51.

Mears, L., S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey
2017a. Application of a mechanistic model as a tool for on-line monitoring of pilot scale filamentous fungal fermentation processes—The importance of evaporation effects. *Biotechnology and Bioengineering*, 114(3):589–599.

Mears, L., S. M. Stocks, M. O. Albaek, G. Sin, and K. V. Gernaey
2017b. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends in Biotechnology*, 35(10):914–924.

Michalowski, A., M. Siemann-Herzberg, and R. Takors
2017. Escherichia coli HGT: Engineered for high glucose throughput even under slowly growing or resting conditions. *Metabolic Engineering*, 40:93–103.

Mockus, J., W. Eddy, A. Mockus, L. Mockus, and G. Reklaitis
1997. *Bayesian Heuristic Approach to Discrete and Global Optimization*. Springer US.

Monod, J.
1949. The Growth of Bacterial Cultures. *Annual Review of Microbiology*, 3(1):371–394.

Monteagudo, J. M., L. Rodríguez, J. Rincón, and J. Fuertes
1997. Kinetics of lactic acid fermentation by Lactobacillus delbrueckii grown on beet molasses. *Journal of Chemical Technology and Biotechnology*, 68(3):271–276.

Morchain, J., J.-C. Gabelle, and A. Cockx
2013. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. *AIChE Journal*, 59(2):369–379.

Moullec, Y. L., C. Gentric, O. Potier, and J. P. Leclerc
2010. Comparison of systemic, compartmental and CFD modelling approaches: Application to the simulation of a biological reactor of wastewater treatment. *Chemical Engineering Science*, 65(1):343–350.

Musvoto, E., M. Wentzel, R. Loewenthal, and G. Ekama
2000. Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems. *Water Research*, 34(6):1857–1867.

Nandasana, A. D. and S. Kumar
2008. Kinetic modeling of lactic acid production from molasses using

Enterococcus faecalis RKY1. *Biochemical Engineering Journal*, 38(3):277–284.

Nauha, E. K., Z. Kálal, J. M. Ali, and V. Alopaeus
2018. Compartmental modeling of large stirred tank bioreactors with high gas volume fractions. *Chemical Engineering Journal*, 334:2319–2334.

Nelsen, R. B.
2006. *An introduction to copulas*. Springer.

Neubauer, P., N. Cruz, F. Glauche, S. Junne, A. Knepper, and M. Raven
2013. Consistent development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*, 13(3):224–238.

Neubauer, P. and S. Junne
2010. Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Current Opinion in Biotechnology*, 21(1):114–121.

Neubauer, P. and S. Junne
2016. Scale-Up and Scale-Down Methodologies for Bioreactors. In *Bioreactors*, Pp. 323–354. Wiley-VCH Verlag GmbH & Co. KGaA.

Nienow, A. W.
2006. Reactor Engineering in Large Scale Animal Cell Culture. *Cytotechnology*, 50(1-3):9–33.

Nienow, A. W.
2014. Re "Development of a scale-down model of hydrodynamic stress to study the performance of an industrial CHO cell line under simulated production scale bioreactor conditions" [Sieck, J.B., Cordes, T., Budach, W.E., Rhi. *Journal of Biotechnology*, 171:82–84.

Ödman, P., C. L. Johansen, L. Olsson, K. V. Gernaey, and A. E. Lantz
2009. On-line estimation of biomass, glucose and ethanol in Saccharomyces cerevisiae cultivations using in-situ multi-wavelength fluorescence and software sensors. *Journal of Biotechnology*, 144(2):102–112.

Ohara, H., K. Hiyama, and T. Yoshida
1992. Kinetic study on pH dependence of growth and death of Streptococcus faecalis. *Applied Microbiology and Biotechnology*, 38(3):403–407.

Oliveira, A. P., J. Nielsen, and J. Förster
2005. Modeling Lactococcus lactis using a genome-scale flux model. *BMC microbiology*, 5(1):39.

Osman, J. J., J. Birch, and J. Varley
2002. The response of GS-NS0 myeloma cells to single and multiple pH perturbations. *Biotechnology and Bioengineering*, 79(4):398–407.

Pais, D. A. M., M. J. T. Carrondo, P. M. Alves, and A. P. Teixeira
2014. Towards real-time monitoring of therapeutic protein quality in mammalian cell processes. *Current Opinion in Biotechnology*, 30:161–167.

Paul, E. L., V. A. Atiemo-Obeng, and S. M. Kresta, eds.
2003. *Handbook of Industrial Mixing*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Payot, T., M. Fick, and G. Chimique
1997. On-line estimation of lactic acid concentration by conductivity measurement in fermentation broth. *Biotechnology Techniques*, 11(1):17–20.

Peng, R. Y., T. C. K. Yang, H.-e. Wang, Y.-c. Lin, and C. Cheng
1997. Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model. *Journal of the Chinese Agricultural Chemical Society*, 35(5):485–494.

Peter, E. and H. Röck
2012. On-line estimation of lactic acid concentration during batch fermentations of Streptococcus thermophilus based on pH measurement. *12th International Conference on Control, Automation and Systems*, Pp. 851–855.

Peterson, J. J. and K. Lief
2010. The ICH Q8 Definition of Design Space: A Comparison of the Overlapping Means and the Bayesian Predictive Approaches. *Statistics in Biopharmaceutical Research*, 2(2):249–259.

Pigou, M. and J. Morchain
2015. Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. *Chemical Engineering Science*, 126:267–282.

Pigou, M., J. Morchain, P. Fede, M.-I. Penet, and G. Laronze
2017. An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors. *Chemical Engineering Science*, 171:218–232.

Pinelli, D., Y. R. González-Vara, D. Matteuzzi, and F. Magelli
1997. Assessment of kinetic models for the production of L- and D-lactic acid isomers by Lactobacillus casei DMS 20011 and Lactobacillus coryniformis DMS 20004 in continuous fermentation. *Journal of Fermentation and Bioengineering*, 83(2):209–212.

Posch, A. E., C. Herwig, and O. Spadiut
2013. Science-based bioprocess design for filamentous fungi. *Trends in Biotechnology*, 31(1):37–44.

Process Development Forum
2014. Mechanistic Modeling: Does it Have a Future in Process Development?

Prunescu, R. M., M. Blanke, J. G. Jakobsen, and G. Sin
2015. Dynamic modeling and validation of a biomass hydrothermal pretreatment process-a demonstration scale study. *AIChE Journal*, 61(12):4235–4250.

Rantanen, J. and J. Khinast
2015. The Future of Pharmaceutical Manufacturing Sciences. *Journal of Pharmaceutical Sciences*, 104(11):3612–3638.

Rehman, U., W. Audenaert, Y. Amerlinck, T. Maere, M. Arnaldos, and I. Nopens
2017. How well-mixed is well mixed? Hydrodynamic-biokinetic model integration in an aerated tank of a full-scale water resource recovery facility. *Water Science and Technology*, 76(8):1950–1965.

Rigopoulos, S. and A. Jones
2003. A hybrid CFD—reaction engineering framework for multiphase reactor modelling: basic concept and application to bubble column reactors. *Chemical Engineering Science*, 58(14):3077–3089.

Roels, J. A.
1981. THE APPLICATION OF MACROSCOPIC PRINCIPLES TO MICROBIAL METABOLISM. *Annals of the New York Academy of Sciences*, 369(1 Biochemical E):113–134.

Rogers, P. L., L. Bramall, and I. J. McDonald
1978. Kinetic analysis of batch and continuous culture of Streptococcus cremoris HP1. *Can J Microbiol*, 24(4):372–380.

Sagmeister, P., M. Kment, P. Wechselberger, A. Meitz, T. Langemann, and C. Herwig
2013. Soft-sensor assisted dynamic investigation of mixed feed bioprocesses. *Process Biochemistry*, 48(12):1839–1847.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola
2008. *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd.

Samstag, R. W., J. J. Ducoste, A. Griborio, I. Nopens, D. J. Batstone, J. D. Wicks, S. Saunders, E. A. Wicklein, G. Kenny, and J. Laurent
2016. CFD for wastewater treatment: an overview. *Water Science and Technology*, 74(3):549–563.

Sandoval-Basurto, E. A., G. Gosset, F. Bolívar, and O. T. Ramírez
2005. Culture of Escherichia coli under dissolved oxygen gradients simulated in a two-compartment scale-down system: Metabolic response and production of recombinant protein. *Biotechnology and Bioengineering*, 89(4):453–463.

Sawatari, Y. and A. Yokota
2007. Diversity and Mechanisms of Alkali Tolerance in Lactobacilli. *Applied and Environmental Microbiology*, 73(12):3909–3915.

Schepers, A. W., J. Thibault, and C. Lacroix
2002a. Lactobacillus helveticus growth and lactic acid production during pH- controlled batch cultures in whey permeate / yeast extract medium . Part I . multiple factor kinetic analysis. *Enzyme and Microbial Technology*, 30:176–186.

Schepers, A. W., J. Thibault, and C. Lacroix
2002b. Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. *Enzyme and Microbial Technology*, 30(2):187–194.

Schmalzriedt, S., M. Jenne, K. Mauch, and M. Reuss
2003. *Integration of Physiology and Fluid Dynamics*, Pp. 19–68. Berlin, Heidelberg: Springer Berlin Heidelberg.

Seber, G. A. F. and C. J. Wild
1989. Nonlinear Regression. (1):792.

Simen, J. D., M. Löffler, G. Jäger, K. Schäferhoff, A. Freund, J. Matthes, J. Müller, and R. Takors
2017. Transcriptional response of Escherichia coli to ammonia and glucose fluctuations. *Microbial Biotechnology*, 10(4):858–872.

Sin, G. and K. V. Gernaey
2016. Data Handling and Parameter Estimation. In *Experimental Methods in Wastewater Treatment*, M. C. van Loosdrecht, P. H. Nielsen, C. Lopez-Vazquez, and D. Brdjanovic, eds., Pp. 201–234. IWA Publishing Company.

Sin, G., K. V. Gernaey, and A. E. Lantz
2009a. Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. *Biotechnology progress*, 25(4):1043–1053.

Sin, G., K. V. Gernaey, M. B. Neumann, M. C. M. van Loosdrecht, and W. Gujer
2009b. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, 43(11):2894–2906.

Sin, G., A. S. Meyer, and K. V. Gernaey
2010. Assessing reliability of cellulose hydrolysis models to support biofuel process design—Identifiability and uncertainty analysis. *Computers & Chemical Engineering*, 34(9):1385–1392.

Sin, G., P. Ödman, N. Petersen, A. E. Lantz, and K. V. Gernaey
2008. Matrix notation for efficient development offirst-principlesmodels within PAT applications: Integrated modeling of antibiotic production with-Streptomyces coelicolor. *Biotechnology and Bioengineering*, 101(1):153–171.

Sobol', I. M.
1967. Point distribution in a cube and approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki I Matematicheskoi Fiziki*, 7(4).

Soini, J., C. Falschlehner, C. Liedert, J. Bernhardt, J. Vuoristo, and P. Neubauer
2008. Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W3110. *Microbial Cell Factories*, 7(1):30.

Solle, D., B. Hitzmann, C. Herwig, M. P. Remelhe, S. Ulonska, L. Wuerth, A. Prata, and T. Steckenreiter
2017. Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. *Chemie Ingenieur Technik*, 89(5):542–561.

Stocker, E., G. Toschkoff, S. Sacher, and J. G. Khinast
2014. Use of mechanistic simulations as a quantitative risk-ranking tool within the quality by design framework. *International Journal of Pharmaceutics*, 475(1-2):245–255.

Stocks, S. M.
2013. Industrial enzyme production for the food and beverage industries: process scale up and scale down. In *Microbial Production of Food Ingredients, Enzymes and Nutraceuticals*, Pp. 144–172. Elsevier.

Sunya, S., C. Bideaux, C. Molina-Jouve, and N. Gorret
2013. Short-term dynamic behavior of Escherichia coli in response to successive glucose pulses on glucose-limited chemostat cultures. *Journal of Biotechnology*, 164(4):531–542.

Tabora, J. E. and N. Domagalski
2017. Multivariate Analysis and Statistics in Pharmaceutical Process Research and Development. *Annual Review of Chemical and Biomolecular Engineering*, 8(1):403–426.

Toumi, A., C. Jürgens, C. Jungo, B. A. Maier, V. Papavasileiou, and D. P. Petrides
2010. Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools. *Pharmaceutical Engineering*, 30(2):1–9.

Venkatesh, K., M. Okos, and P. Wankat
1993. Kinetic model of growth and lactic acid production from lactose by Lactobacillus bulgaricus. *Process Biochemistry*, 28(4):231–241.

Villadsen, J., J. Nielsen, and G. Lidén
2011. *Bioreaction Engineering Principles*. Boston, MA: Springer US.

Vlaev, D., R. Mann, V. Lossev, S. D. Vlaev, J. Zahradnik, and P. Seichter
2000. Macro-mixing and streptomyces fradiae modelling oxygen and nutrient segregation in an industrial bioreactor. *Chemical Engineering Research and Design*, 78(3):354–362.

von Stosch, M., R. Oliveira, J. Peres, and S. F. de Azevedo
2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, 60:86–101.

Vrabel, P., R. G. J. M. Van Der Lans, K. C. A. M. Luyben, L. Boon, and A. W. Nienow
2000. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: Modelling and measurements. *Chemical Engineering Science*, 55(23):5881–5896.

Vrábel, P., R. G. J. M. van der Lans, F. N. van der Schot, K. C. Luyben, B. Xu, and S.-O. Enfors
2001. CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chemical Engineering Journal*, 84(3):463–474.

Wang, G., J. Chu, H. Noorman, J. Xia, W. Tang, Y. Zhuang, and S. Zhang
2014. Prelude to rational scale-up of penicillin production: a scale-down study. *Applied Microbiology and Biotechnology*, 98(6):2359–2369.

Wechselberger, P., A. Seifert, and C. Herwig
2010. PAT method to gather bioprocess parameters in real-time using simple input variables and first principle relationships. *Chemical Engineering Science*, 65(21):5734–5746.

Wells, G. J. and W. H. Ray
2005. Methodology for modeling detailed imperfect mixing effects in complex reactors. *AIChE Journal*, 51(5):1508–1520.

Whitman, W. G.
1962. The two film theory of gas absorption. *International Journal of Heat and Mass Transfer*, 5(5):429–433.

Xu, B., M. Jahic, G. Blomsten, and S.-O. Enfors
1999. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli. *Applied Microbiology and Biotechnology*, 51(5):564–571.

Yang, Y., J. Xia, J. Li, J. Chu, L. Li, Y. Wang, Y. Zhuang, and S. Zhang
2012. A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *Journal of Biotechnology*, 161(3):250–256.

Youssef, C. B., G. Goma, and A. Olmos-Dichara
2005. Kinetic modelling of Lactobacillus casei ssp. rhamnosus growth and lactic acid production in batch cultures under various medium conditions. *Biotechnology Letters*, 27:1785–1789.

Zacharof, M.-P. P. and R. W. Lovitt
2013. Modelling and simulation of cell growth dynamics, substrate consumption, and lactic acid production kinetics of Lactococcus lactis. *Biotechnology and Bioprocess Engineering*, 18(1):52–64.

Zahradník, J., R. Mann, M. Fialová, D. Vlaev, S. D. Vlaev, V. Lossev, and P. Seichter
2001. A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. *Chemical Engineering Science*, 56(2):485–492.

Zavatti, V., H. Budman, R. Legge, and M. Tamer
2016. Monitoring of an antigen manufacturing process. *Bioprocess and Biosystems Engineering*, 39(6):855–869.

Zhao, L., H.-Y. Fu, W. Zhou, and W.-S. Hu
2015. Advances in process monitoring tools for cell culture bioprocesses. *Engineering in Life Sciences*, 15(5):459–468.

Zhao, W., A. Buffo, V. Alopaeus, B. Han, and M. Louhi-Kultanen
2017. Application of the compartmental model to the gas-liquid precipitation of CO2-Ca(OH)2aqueous system in a stirred tank. *AIChE Journal*, 63(1):378–386.

Zou, X., J.-y. Xia, J. Chu, Y.-p. Zhuang, and S.-l. Zhang
2012. Real-time fluid dynamics investigation and physiological response for erythromycin fermentation scale-up from 50 L to 132 m3 fermenter. *Bioprocess and Biosystems Engineering*, 35(5):789–800.

Part II
Journal Articles and Peer-Reviewed Conference Proceedings

Paper A

Modelling for Process Risk Assessment in Industrial Bioprocesses

Robert Spann[a], Anna Eliasson Lantz[b], Krist V. Gernaey[a], Gürkan Sin[a]

[a] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark

Published in the Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Elsevier, 2018.
https://doi.org/10.1016/B978-0-12-409547-2.14356-2

Abstract

The objective of this contribution is to give an overview about model implementations for risk-based decision making in industrial bioprocesses, such as the antibiotic production for the pharmaceutical industry. It focuses on the applications of mechanistic and computational fluid dynamics (CFD) models. The models are built to support the understanding of the process, improve or speed up the process development, and used to monitor and control the production process to achieve the desired product quality and quantity. Uncertainties inherently present in development of these models are considered in many applications, e.g. by for example using Monte Carlo simulations, to enable a risk-based decision making when the model results from Monte Carlo simulations are assessed. The sources of uncertainties may include for example process input variations, model parameter uncertainty, assumptions underlying the model structure, and measurement errors, among others. More and more studies combine mechanistic biochemical models with CFD models to investigate especially heterogeneous process conditions at large scale such as substrate gradients. However, on-line applications of CFD models, e.g. for process control, are hampered by the long computation times. Instead, CFD modelling efforts are directed towards supporting CFD-based compartment modelling that reduces the spatial resolution, but allows a much faster simulation compared to a CFD model. Compartment models integrating bio-kinetics of the bioprocess, various sources of uncertainties in the system, and heterogeneities in the bioreactor can then be applied as an enabling tool for risk-based on-line monitoring and control systems to achieve optimized bioprocess operations.

Glossary

- Computational fluid dynamics (CFD): Modelling approach to model fluid flows in systems.
- Critical process parameter (CPP): “A process parameter whose variability has an impact on a critical quality attribute and therefore should be monitored or controlled to ensure the process produces the desired quality.” (ICH Q8(R2), 2009)
- Critical quality attributes (CQAs): “A physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality.” (ICH Q8(R2), 2009)
- Design space: “The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality.” (ICH Q8(R2), 2009)
- Mechanistic model: A model based on first-principles using the fundamental bioenergetics principles and biochemical reactions.
- Off-line: antonym of on-line (see below).
- On-line: real time. There is a further nuance when considering on-line measurements in comparison to at-, in-, and off-line measurements. (FDA, 2004)
- Process analytical technology (PAT): Utilization of tools to design, analyse, and control bioprocesses through the measurement and/or prediction of critical process parameters. (FDA, 2004)
- Quality by design (QbD): “A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management.” (ICH Q8(R2), 2009)

Introduction

Microbial cultivations are frequently used to produce various products e.g. for the food, chemical, and pharmaceutical industry. Microorganisms, such as bacteria, yeast, fungi, and mammalian cells are cultivated in large-scale bioreactors where the microorganisms grow and convert a substrate into the desired end product. The bioreactors have a volume of up to several 100 m^3. It is of high importance to assure a high product quality in order to guarantee the right effect e.g. of a pharmaceutical drug as well as a cost effective production. It is challenging to control a manufacturing process in such a way that the desired product quality is achieved because of the large natural variability in large-scale microbial processes. The production process is subject to uncertainties since process inputs, e.g. the living microorganisms or raw materials, vary in each production cycle. Running a production with identical settings does therefore not automatically lead to a constant product quality. It is necessary to control the bioprocess dynamically and to counteract unwanted changes. However, there is a lack of real-time measurement techniques that allow to track the product quality attributes in a microbial cultivation process in real time (Pais, Carrondo, Alves, & Teixeira, 2014), hence the product quality attributes have to be predicted indirectly by appropriate models or soft sensors.

The process analytical technology (PAT) guidance of the US Food and drug administration (FDA, 2004) aims for improved monitoring and control systems to ensure product quality in bioprocesses. “The goal of PAT is to enhance understanding and control the manufacturing process” (FDA, 2004). A strategy to ensure product quality in the bioprocess is the quality by design (QbD) approach. There, a design space, i.e. operation space, for the production process is defined rather than using fixed parameter settings. The desired product quality is thereby ensured in each production cycle (Rathore & Winkle, 2009). The product quality profile is defined by critical quality attributes (CQAs) of the product, such as stability, purity, density, size, concentration (ICH Q8(R2), 2009). Since CQAs are typically properties and characteristics that can hardly be measured in real time, they are correlated with on-line measureable process variables, such as temperature, pH, O_2 concentration, spectroscopic measurements (e.g. near-infrared (NIR), Raman), etc. to define the design space. In the QbD approach, critical process parameters (CPPs), i.e. process parameters that have an impact on the CQAs, are determined. The effect of the CPPs on the CQAs is finally examined by applying statistical analyses (Mercier, Diepenbroek, Wijffels, & Streefland, 2014; Rathore, Mittal, Pathak, & Arora, 2014) or mechanistic modelling (Sommeregger et al., 2017). More specifically, a model is used to predict the unmeasurable CQAs of the product by using the measured CPPs as model inputs while the process is running. The CPPs are then controlled within the design space during the production so that the target values for the product CQAs will be achieved.

Mechanistic and data-driven models are therefore becoming increasingly important for biotechnological processes and find many applications especially as PAT systems (Glassey et al., 2011; Koutinas, Kiparissides, Pistikopoulos, & Mantalaris, 2012; Rantanen & Khinast, 2015). Mechanistic models describe the complex biological, chemical, and physical phenomena in the bioreactor using mechanistic understanding of the process. Data-driven models are trained with large data sets, and sometimes difficult to interpret because their algorithms have no physical meaning, but they are nevertheless frequently applied (Djuris & Djuric, 2017). To fulfill the QbD approach, the PAT framework involves many activities, from development and installation of novel analytical sensors for measurement solutions, over the implementation of risk assessment strategies, to advanced data analysis methods. This includes the broad view of the system reflecting chemical, physical and microbial measurements. Dynamic models are developed for enabling process systems engineering strategies in the context of PAT such as the design, understanding, optimization, monitoring, and control of the manufacturing

processes. Even though the PAT guidelines were initially addressed to the pharmaceutical industry, many other life sciences industries, such as industrial biotechnology that includes the food and feed industry, are implementing this methodology, as well.

Large-scale bioprocesses are usually hampered by suboptimal process conditions due to the heterogeneous conditions in large bioreactors resulting from lower mixing capabilities compared to the mixing in lab-scale reactors. That often causes reduced yields or product titers (Lara, Galindo, Ramírez, & Palomares, 2006). Modelling of large-scale microbial cultivations therefore demands both the simulation of the mixing behaviour (fluid dynamics) and the biochemical system. Computational fluid dynamics (CFD) simulations are applied to solve the fluid dynamics of the bioreactors, and they are then combined with biokinetic modelling to support development of process understanding and systems engineering for optimization and control. The disadvantage is, however, that CFD simulations require many computational resources, and are not applicable for on-line applications as it would be required for risk-based on-line monitoring and control. To this end, we see the development of process systems engineering towards CFD-based compartment models that employ the information from the CFD simulations into model-based applications, albeit at a reduced spatial resolution as an attractive alternative (Box 1). Compartment models allow much faster simulations.

Box 1 Bioprocess modelling for improving bioprocess operations.

Mechanistic models are increasingly applied in biotechnological processes for various process analytical technology (PAT) systems. For large-scale process modelling, both the biological phase and the fluid dynamics need to be considered, since conditions are not homogeneously distributed at large scale. Instead, gradients of e.g. the substrate concentration exist in fed-batch or continuous cultivations. Computational fluid dynamic (CFD) models and compartment models are applied to take the heterogeneities into account. Furthermore, there are uncertainties in the cultivation and modelling system that need to be accounted for in order to allow a risk-based decision making. Uncertainties may include process input variations, model parameter uncertainties, and assumptions underlying the model structure. This approach then allows a risk-based monitoring and control of processes, and hence improves both the performance and the safety of the production process.

In risked-based bioprocess model applications for large scale, both a mechanistic bioprocess model that describes the biochemical processes, and a CFD model that describes the fluid dynamics of the bioreactor are developed initially (Fig. 1, 1st row). On the one hand, the mechanistic model can be applied as a single-compartment model, which is the typical approach used in the past for fermentation modelling studies. On the other hand, the mechanistic model can be combined with a multi-compartment model that is derived from the CFD simulation (Fig. 1, 2nd row). Both modelling approaches are suitable for many applications. Production processes can for example be developed and tested off-line, or monitored and controlled on-line (Fig. 1, 3rd row). Performing an uncertainty analysis, e.g. with a Monte Carlo simulation, allows then the plant operators to make decisions based on a risk assessment. The aim of this methodology is to achieve an optimal bioprocess (Fig. 1, bottom row). The sub-steps may need repetitions, as this methodology has to be understood as an iterative procedure to improve the model and application quality.

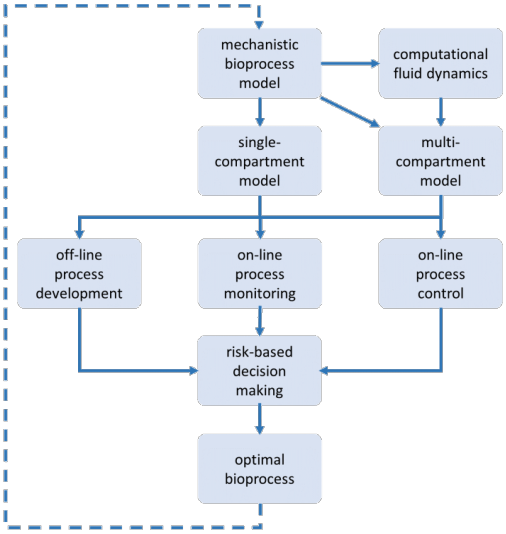


FIGURE 1 Flow diagram for risk-based bioprocess model applications.

This contribution describes model implementations for process risk assessment in industrial bioprocesses. It focuses on applications of process systems engineering tools for large-scale production processes, where complex biological, chemical, and physical phenomena are interconnected and simulated by – mainly mechanistic – models. First, an overview of the production process of fine chemicals and pharmaceutical products by microbial cultivations is given. Second, state of the art applications for mechanistic models of microbial, chemical and physical phenomena in industrial bioprocess are described. Third, probabilistic modelling approaches and risk assessment considering natural uncertainties and the probability of process failures in the model description are summarized. Then, heterogeneities, which occur at large-scale cultivations, are considered and kinetic models integrated into computational fluid dynamics and compartment models are demonstrated. These tools contribute to expanding the process systems engineering toolbox for on-line risk-based monitoring and control applications at large-scale production processes. Finally, future directions of the PAT framework are outlined.

Bio-production of chemicals

Many biotechnological products are manufactured by means of large-scale cultivations. The products range from bulk chemicals (e.g. ethanol, acetone, butanol, organic acids), over fine chemicals (e.g. enzymes, polymers) to pharmaceuticals (e.g. antibiotics, antibodies, steroids). Low-value products have usually a high market volume, while high value products are sold in low quantities (Junker, 2004; Nielsen, 2003). Cells produce the products in bioreactors. The process must be kept aseptic for many products, since contaminations could lead to the loss of the production. The cell factories comprises many species of bacterial, yeast, fungal, plant, animal, mammalian, and stem cells. While some processes use wild-type or natural mutant strains, genetically modified microorganisms are increasingly

applied to meet the needs of the growing demand of a sustainable production (Campbell, Xia, & Nielsen, 2017; Meyer, Minas, & Schmidhalter, 2016).

A bioreactor is designed for optimal gas, heat, and mass transfer ensuring a short mixing time. However, some cells, such as mammalian cells, are shear-sensitive, and a compromise between fast mixing and low shear stress must be made. The most commonly used bioreactor is the stirred tank bioreactor (STR). It is a vessel made out of stainless steel consisting of a stirring system (motor connected to a shaft with one or more impellers) to mix the cultivation broth, baffles to enhance mass transfer and avoid the formation of a vortex, and a cooling system to cool down the liquid since the biological metabolism releases heat. The working volumes of the bioreactor depend on the product category ranging from cheap commodities to high value products. Bulk chemicals, or animal feed supplements, such as amino acids and vitamins, are produced in bioreactors with working volumes of up to several hundred cubic meters. Pharmaceuticals are typically produced with working volumes of several tens of cubic meters. There are many other bioreactor configurations but it is outside the scope of this contribution to address all of them (for review on this topic, see Chisti & Moo-Young (2003); Mandenius (2016)).

There are three fundamental operation modes for microbial cultivations: the batch cultivation, the fed-batch cultivation, and the continuous cultivation (Box 2).

Box 2 Bioprocess operation modes.

Microbial fermentations are usually performed either in batch, fed-batch, or continuous mode. In the **batch** process, the medium containing all nutrients is already available in the bioreactor from the beginning (Fig. 2, left). The only components that might be added to the process during the operation are gasses (O_2, CO_2, N_2), a base or acid for pH control, an antifoaming agent, or an inducer for gene expression. The process is finished when one component, e.g. the substrate, is limiting, and the production stops. In a **fed-batch** process, a concentrated solution of medium components, which is often the carbon substrate and a nitrogen source, are continuously added via an inlet to the process while it is running (Fig. 2, middle). In the **continuous** process, all nutrients are continuously added to the process and the cultivation broth is withdrawn from the bioreactor at the same time to keep the bioreactor volume constant (Fig. 2, right).

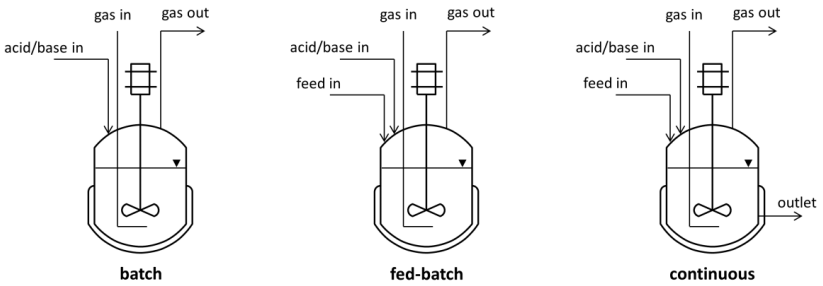


FIGURE 2 Bioprocess operation modes: batch, fed-batch and continuous cultivation.

Mechanistic models for industrial bioprocesses

A mechanistic model describes the behaviour of a system with mathematical equations. Mechanistic models are applied to develop, optimize, control, etc. different bioprocesses. They are based on prior knowledge of the phenomena of the system, which are in particular the elemental mass, energy, and momentum balances in bioprocesses (Esener, Roels, & Kossen, 1983; Roels, 1981). Mechanistic models are therefore also called first principles, fundamental, or white-box models. Usually ordinary differential equations (ODE) describe the system, if the model outputs change dynamically in time. The differential equations for all system components cover typically biological, chemical, and physical mechanisms, such as microbial growth, pH calculation, and aeration, respectively (see section Biological, chemical, and physical model expressions for cultivations). The main components are for example the substrate (carbon source), biomass, oxygen, reactor volume, and gas flows in an aerobic cultivation.

The main advantage of mechanistic models is that they have a large extrapolation capability, since they are based on first principles, and are not limited to the conditions that were used to calibrate the model. Mechanistic models have therefore a large application range (see section Applications of process models) and can be used to test scenarios even if no data of these conditions are available (Mears, Stocks, Albaek, Sin, & Gernaey, 2017b). Another advantage of mechanistic models is that the model parameters have a physiological meaning, e.g. the specific maximum growth rate of the cells; hence, the user of the model can directly understand them. Furthermore, such models have a general and flexible structure that allows knowledge transfer of process equations and parameters both in the industrial and academic environment from one process model to another. However, disadvantages of mechanistic models are that they require a relatively long development time, and that a significant process insight is needed to obtain and validate the fundamental process equations.

In the end, mechanistic models for the bio-based production industry are often a combination of mechanistic and empirical equations, due to the complexity of the system. The latter is used especially for situations, where some phenomena are not understood to the necessary level of detail. There are also several other types of models for bioprocesses that are out of the scope of this contribution, including data-driven models (which include for example artificial neural networks and chemometric methods, such as principal component regression), and hybrid models (a combination of mechanistic and data-driven modelling techniques (Solle et al., 2017; von Stosch, Oliveira, Peres, & Feyo de Azevedo, 2014).

The biotechnological industry increasingly applies mechanistic models, because the industry has realized the significance of mechanistic models. Mechanistic models are for example advantageous to predict the system behaviour after a change, while statistical models based on QbD sometimes fail (Process Development Forum, 2014). Today data-driven approaches are preferred in industry because of the cost-benefit analysis, which might be owing to the established regulatory requirements for the approval of new processes. However, mechanistic models can support the PAT framework to better understand the effect of process changes under a wider range of conditions. Mechanistic models can simulate different initial conditions as well as process disturbances thanks to their extrapolation capabilities. They use the critical process parameters (CPPs) as input to predict the critical quality attributes (CQAs) e.g. of the desired product.

Once the model structure is established, the model parameters have to be determined and the reliability of the model must be assessed. It is important to prove the reliability of the model in order to be able to rely on the model-based application, and to be able to defend for example the design space towards the

FDA. Several methods and tools exists to assess the credibility of the model including identifiability, uncertainty and sensitivity analysis (Chis, Banga, & Balsa-Canto, 2011; Sin & Gernaey, 2016).

Biological, chemical, and physical model expressions for cultivations

The differential model equations typically cover biological, chemical and physical processes. Biological processes are often modelled with a macroscopic view, describing the microbial cell as a whole (unstructured model), especially when the model is applied for the large-scale process design and optimization (Gernaey, Lantz, Tufvesson, Woodley, & Sin, 2010; Kroll, Hofer, Stelzer, & Herwig, 2017). In structured models, at least one intracellular metabolic component, such as metabolites, ATP, and NADH, are described. This is of particular interest for metabolic engineering and microbial cell development in system biology (Almquist, Cvijovic, Hatzimanikatis, Nielsen, & Jirstrand, 2014; Campbell et al., 2017), which is outside the scope of this contribution. In the unstructured biological model, biomass growth, product formation, maintenance, and decay processes (of biomass, products, and precursors) are described. The growth of biomass is written with the stoichiometric equation (Eq. (1)), where biomass ($CH_aO_bN_cP_d$), a product (P), and carbon dioxide (CO_2) are typically obtained from the conversion of the substrate (e.g. Glucose CH_2O), oxygen (O_2), and a nitrogen (e.g. NH_3) and phosphate source (e.g. H_3PO_4) (Villadsen, Nielsen, & Lidén, 2011). Biomass is composed of many elements: C, H, O, N, P, S, and trace elements such as Ca and Mg, which can be analyzed by an elemental analysis. Only the dominant elements are usually considered for modelling purposes, and a pseudo steady-state of this composition is assumed, even though the elemental composition of the cell might slightly change during the process. The yield coefficients define the quantity of something produced/consumed per quantity of consumed substrate, e.g. the biomass yield coefficient Y_{SX} that describes how much biomass is produced per consumed substrate. The kinetic rates of the biological processes have been described in the literature using empirical relations, such as the Monod model (Monod, 1949) combined with functions accounting for inhibition and limiting effects of substrates, metabolites, products, and process conditions like the pH (Eq. (2)).

$$1CH_2O + Y_{SO}O_2 + Y_{SN}NH_3 + Y_{SPhos}H_3PO_4 \rightarrow Y_{SX}CH_aO_bN_cP_d + Y_{SP}P + Y_{SC}CO_2 + (\cdots) + Y_{SW}H_2O \quad (1)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot \frac{C_S}{C_S + K_s} \cdot \frac{K_P}{C_P + K_P} \cdot e^{-\left(\frac{(|pH_{opt} - pH|)^2}{\sigma^2}\right)} \cdot C_X \quad (2)$$

where dC_X/dt is the change of the biomass concentration (C_X) over time (t), C_S is the substrate concentration, K_S the saturation parameter of the substrate, C_P the product concentration, K_P the product inhibition parameter, pH_{opt} the optimal pH for growth, σ the spread parameter of the Gaussian pH function. The latter approximates a bell like curve relating the pH effect on the maximum growth of cells.

Chemical processes are considered in bioprocess modelling in order to extent the application of models to systems e.g. with varying pH. These models are applied to study for example the effect of the pH or precipitation on the process. Several chemical processes are modelled including dissociation reactions of weak and strong acids and bases, ion pairing, and precipitation (Musvoto, Wentzel, Loewenthal, & Ekama, 2000). Ion pairing and precipitation involve solid-liquid interactions, while a mixed acid/base system describes liquid-liquid interactions. In these processes, the analytical concentrations need to be adjusted by the activity coefficients (γ_i), due to the changing interactions of ions in solution with each other and with the H_2O molecules at different ion concentrations. Different empirical equations are applied for such activity corrections, depending on the ionic strength (function of the concentration of all charged components) of the mixture (Loewenthal, Ekama, & Marais, 1989; Musvoto et al., 2000).

As an example, Eq. (3) shows the dissociation reaction rate of an undissociated acid (HA) to its dissociated form (A^-) and the hydrogen ion (H^+):

$$r_{HA \rightleftharpoons A^- + H^+} = K'_{r,A} \cdot 10^{-pK_A/\gamma_A^2} \cdot C_{HA} - K'_{r,A} \cdot C_{A^-} \cdot C_{H^+} \quad (3)$$

where $K'_{r,A}$ is the apparent reverse rate constant for the acid dissociation, the pK_A is the acid dissociation constant for the specific acid considered, and C_i represents the concentration of the components.

Physical processes deal with mass and heat transfer processes. Gas-liquid exchange processes are of special interest in cultivations. These include among others the mass transfer of the oxygen from the gas bubbles to the liquid broth (aeration) and the gas stripping of CO_2 from the liquid to the gas phase. Oxygen transfer is key for the success of most of the large-scale bioprocesses, since most of them are aerobic – with the exception of e.g. bioethanol and lactic acid production. Gas-liquid mass transfer models are usually based on the two-film theory (Whitman, 1923), and can be described as the product of the volumetric mass transfer coefficient, k_La, and the driving force (concentration difference of the component in the liquid phase (C_{O2}), and the gas-liquid interface(C^*_{O2})) (Eq. (4)). There are numerous empirical correlations for k_La (Markopoulos, Christofi, & Katsinaris, 2007), and several experimental methods have been developed to measure the gas-liquid mass transfer (Villadsen et al., 2011).

$$r_{O_2(g) \rightleftharpoons O_2(l)} = k_L a_{O_2} \left(C^*_{O_2} - C_{O_2} \right) \quad (4)$$

Complex, non-linear models incorporate various biological, chemical, and physical processes. Detailed models include countless processes in order to better understand for example the anaerobic digestion process (Flores-Alsina et al., 2016), the pharmaceutical production of penicillin by *Penicillium chrysogenum* (Goldrick, Stefan, Lovett, Montague, & Lennox, 2015), different aeration and agitation conditions for the enzyme production with *Aspergillus oryzae* (Albaek, Gernaey, Hansen, & Stocks, 2011), the antibiotic production with *Streptomyces coelicolor* (Sin, Ödman, Petersen, Lantz, & Gernaey, 2008), and *Pichia pastoris* and *Saccharomyces cerevisiae* fed-batch cultivations (Lencastre Fernandes et al., 2013; Wechselberger, Seifert, & Herwig, 2010). Many mechanistic models rely on a similar model structure, based on first-principles chemical and physical process descriptions that have been added to the description of the biological mechanisms. This is due to the flexible structure of mechanistic models that allows them to be adapted easily to other bioprocesses.

Applications of process models

Mechanistic models can be applied to a range of tasks from the stage of process development to the implementation at the industrial-scale production (Mears et al., 2017b): They are applied for example to process development, optimization, monitoring, and control. The applications contribute to an economic and sustainable production and comply also with the PAT guidelines (Gernaey, Cervera-Padrell, & Woodley, 2012).

Mechanistic models are applied in the process development phase to design a new process. Design means typically that a new process is designed for an existing plant since it is rather rare that a new production plant is constructed for a new process. This process involves the definition of optimal process operating conditions, such as stirrer speed, substrate addition rate (for fed-batch cultivations), and aeration rate under the given constraints of the available equipment, as demonstrated for an enzyme production with *Aspergillus oryzae* (Albaek et al., 2011). Initial biotechnological strain and process development is done at lab scale, which is further refined and validated at pilot-scale conditions. The final process is then scaled-up to the production scale because experiments at large scale are rather expensive. Mechanistic models are especially attractive at this stage, since often only very little data is

available, and they are capable of testing new conditions or new equipment for which there is little or no data available. The models can help to assess and understand the effects of the large-scale process conditions, since process scale-up is very challenging (Neubauer et al., 2013; Stocks, 2013; G. Wang et al., 2014). Traditionally, bioprocess scale-up is based on physical parameters, such as volumetric power input, the volumetric oxygen mass transfer coefficient k_La, or stirrer tip speed. However, biological properties and the effect of the heterogeneous cultivation environment (see section Heterogeneities at large-scale cultivations) are often not taken into account. Another challenge are varying process conditions, such as humidity, aeration rate, impeller speed or type, when different bioreactors are used at different production sites all over the world, but the same product quality is of course demanded.

Optimizing an established process is key for the long-lasting success of a manufacturer in order to withstand the economic pressures in a competitive market. Mechanistic models have been successfully applied to optimize cultivations for many processes. As an example, (Jiménez-Hornero, Santos-Dueñas, & García-García, 2009) proposed a model-based optimization for the acetic acid cultivation. Models are also applied to optimize not the main cultivation itself, but the pre-cultures instead. Frahm (2014) optimized the seed train for biopharmaceutical production. The seed is required to inoculate the production bioreactor with the right amount of cells in order to allow a stable production. (Toumi et al., 2010) used multi-unit process simulation in order to optimize a large-scale monoclonal antibody production.

Real-time determination of CQAs is of utmost interest in the biotechnological industry. However, only a few probes are available that measure the required attributes in the bioreactor (in-line). Biomass can for example be measured indirectly in-line by capacitance or turbidity measurements. However, changing conditions of the complex biological process matrix influence these measurement techniques, and make a reliable prediction challenging. For the same reason, probes measuring metabolite concentrations (e.g. product concentrations) are challenging, as well. Further reasons for the current lack of in-line probes are that they must be robust, give stable signals, and withstand harsh conditions during sterilization and cleaning procedures. In addition, they must fulfill the regulatory requirements under the good manufacturing practice (GMP).

Software sensors are an established alternative to hardware probes for the monitoring of bioprocesses and countless implementations in the bioprocess industry can be found (Biechele, Busse, Solle, Scheper, & Reardon, 2015; Luttmann et al., 2012; C. F. Mandenius & Gustavsson, 2015; Pais et al., 2014; Posch, Herwig, & Spadiut, 2013; Sagmeister et al., 2013; Spann et al., 2018; L. Zhao, Fu, Zhou, & Hu, 2015). Software sensors are often also called soft sensors or state estimators. Soft sensors utilize the available on-line measurements, such as the exhaust gas analysis results (e.g. O_2, CO_2, and other volatile compounds), pH, temperature, pressures, flow rates, stirrer speed. These measurements are then used as input for the model that predicts the unknown process states including the CQA. In addition to the above-mentioned measurements, spectroscopic methods are sometimes applied in combination with data driven models (Kadlec, Gabrys, & Strandt, 2009). Spectroscopic methods include UV-Vis (ultraviolet–visible) (Ödman, Johansen, Olsson, Gernaey, & Lantz, 2009; Zavatti, Budman, Legge, & Tamer, 2016), near-infrared (Alves-Rausch, Bienert, Grimm, & Bergmaier, 2014), mid-infrared fluorescence spectroscopy (Fayolle, Picque, & Corrieu, 1997), and Raman spectroscopy (Golabgir & Herwig, 2016), among others. Automated microscopy for acquiring cell-specific information, such as cell morphology, are intensively studied for monitoring (Marquard, Schneider-Barthold, Düsterloh, Scheper, & Lindner, 2017) and control strategies (Bluma et al., 2010), e.g. for the bioethanol production process (Belini, Wiedemann, & Suhr, 2013).

Models are also applied for control strategies in the bioprocess industry. Bioprocesses are often fed-batch processes, where a feed solution is added continuously to the cultivation broth. Models are applied for example to define the feed flow rate, in order to control the substrate concentration in the broth. Since there are no sufficient hardware probes to measure the substrate concentration in real-time, models are used to predict the substrate concentration and to allow better control strategies (Craven, Whelan, & Glennon, 2014; Johnsson, 2015; Mears, Stocks, Albaek, Sin, & Gernaey, 2017a).

Uncertainties in model predictions

Owing to the complexity bioprocesses and mechanistic models, it is necessary to consider uncertainties when a model is applied. There are three sources of uncertainties: i) stochastic uncertainty; ii) input uncertainty; and iii) structural uncertainty (Michael D. McKay, Morrison, & Upton, 1999):

i) The stochastic uncertainty covers the stochastic variabilities that are observed in real processes, for example random failure of equipment that leads to a disturbance in the process. The unforeseeable shutdown of the air sparging for example would result in a shortage of oxygen.
ii) The input uncertainties refer to the lack of perfect knowledge about the model parameters and model inputs. The model parameters rely on available measured data, which are naturally subject to random measurement errors. A parameter value is characterized by a probability distribution around its nominal value, instead of considering a single value. In addition to the uncertain model parameters, model inputs such as initial process conditions, and on-line measurements are uncertain. The initial substrate concentration, for example, varies from batch to batch owing to variations in the medium preparation process, hence a fixed value cannot be expected.
iii) The structural uncertainties are related to the mathematical description of the model. We have typically insufficient knowledge of the bioprocess that would allow a model description including all relevant details of the process. Our applied models are therefore an approximation of the process and based on assumptions.

In order to analyse the effect of the uncertainties, an uncertainty analysis is performed. This is an important element of good modelling practice to ensure a reliable mechanistic model and a robust PAT application (Sin, Gernaey, & Lantz, 2009). In the uncertainty analysis, the uncertainties are propagated to the model outputs and their effect is evaluated. There are some methods for the uncertainty analysis available including the Monte Carlo procedure, differential analysis, response surface methodology, the Fourier amplitude sensitivity test, and the Sobol' variance decomposition (Helton & Davis, 2003; Saltelli et al., 2008). In this contribution, we focus on the Monte Carlo procedure, as this method is widely used and reliable (Helton & Davis, 2003).

Probabilistic model predictions

The uncertainty analysis provides probabilistic information about the performance of the bioprocess. It shows for example what the probability is that the required product quantity or quality will be achieved by considering the above mentioned input uncertainties. Frequently, the Monte Carlo procedure is used to propagate the input uncertainties to the output predictions (Sin, Gernaey, & Lantz, 2009). The Monte Carlo procedure involves three steps: 1) defining the input uncertainties; 2) sampling within the input uncertainty; and 3) Monte Carlo simulation to obtain the model output uncertainty.

In step 1, the model input uncertainties are defined, which is typically performed based on expert opinions. This involves both asking process experts from industry and academia, and consulting

relevant literature. The aim is to obtain the uncertainty for model parameters, initial process conditions, and measurements, e.g. what is the lower and upper bound of these variables. A probability distribution is then assigned to the variables. Often, a uniform distribution, also known as uninformed prior, is assumed when there is lack of observations or data to support the derivation of a specific distribution function such as normal distribution. The latter is usually assumed to describe the distribution function for measurement errors. For the model parameters, the maximum likelihood estimation theory can be used to derive a multivariate normal distribution and the covariance matrix. Such information should then be used for appropriate definition of input model parameter uncertainties (Sin, Meyer, & Gernaey, 2010).

In step 2, random combinations of the model inputs are sampled considering the above defined uncertainties (Fig. 3, Input samples). As an example, the parameters A and B representing model input parameters are considered to be uncertain, and N=1000 input samples are sampled. There exist several sampling methods including random sampling (Meng, 2013), low-discrepancy sequence such as Halton (Halton, 1964) and Sobol (Sobol', 1967) sequences, and stratified sampling such as Latin Hypercube sampling (LHS) (M. D. McKay, Beckman, & Conover, 1979). The sampling method is used to sample N independent inputs from the probability domain [0 1], where N is the number of input samples. Most of the time, parameters and inputs are correlated (see the covariance matrix estimation from the maximum likelihood theory). To preserve the correlation, dependent samples need to be generated, e.g. by applying the multivariate probability distribution copula (Nelsen, 2006) or the Iman-Conover method (Iman & Conover, 1982). The Iman-Conover method employs a rank-based correlation control method to induce the desired correlation between the parameter samples from an independent input space. For those parameters, initial conditions, and measurements that lack correlation information, an identity correlation matrix is used (meaning no correlation is assumed). Finally, the input samples are inverted from the probability domain [0 1] to real values with an inverse cumulative distribution function corresponding to the distribution function for each input.

In step 3, the input uncertainties are propagated using the Monte Carlo procedure to estimate the output uncertainty (Fig. 3, Monte Carlo simulation). The model is thereby simulated N times with each of the above defined input sample sets.

The results are N predictions of the model outputs (Fig. 3, Model outputs). For each model output, the span of the model output prediction indicates the extent of its uncertainty during the cultivation. Inferential statistical analysis such as mean, and 90^{th} and 10^{th} percentiles is applied to assess the results. It is important to realize that the results of the uncertainties can only be interpreted in the analysis boundaries/frame (Sin, Gernaey, Neumann, van Loosdrecht, & Gujer, 2009). This is due to the definitions and assumptions that are made in the study, e.g. the selected model, the identification and characterization of uncertainties, and the selected methodology.

Many studies have investigated the effect of uncertainties on the model outputs for different biotechnological processes, including a hydrothermal pre-treatment process of lignocellulosic biomass (Prunescu, Blanke, Jakobsen, & Sin, 2015), a milk drying process (Ferrari, Gutiérrez, & Sin, 2016), and the antibiotic production with *Streptomyces coelicolor* (Sin et al., 2008).

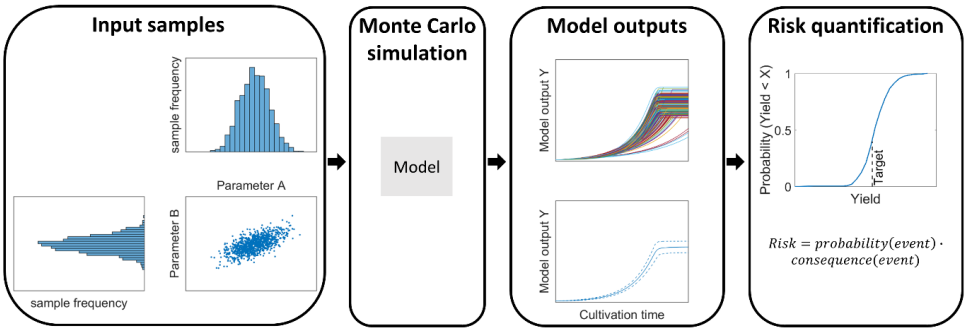


FIGURE 3 Risk quantification using the Monte Carlo procedure with four steps. 1. Input uncertainty space and sampling; 2. Monte Carlo simulation; 3. Model outputs; 4. Risk quantification.

Risk-based decision making

Bioprocesses at production scale undergo a risk assessment to ensure the quality of the manufacturing. The risk assessment aims to increase the safety of the process, and the quality of the product. Most commonly, risk management tools recommended by (ICH Q9, 2005) are used, such as the Failure Mode and Effects Analysis (FMEA). In this worksheet-based method, the risks related to e.g. material properties and process conditions are quantified, first. This characterization is fundamental to understand the impact of critical process parameters (CPPs) on the critical quality attributes (CQAs), most likely the critical quality attributes of the product (see Introduction). Then, the most important critical process parameters are prioritized and further experiments could be carried out, if needed. Finally, the design space for the production is defined, i.e. the acceptable variability e.g. in material properties and process conditions is defined. The process operation within the design space minimizes then the risks of obtaining faulty batches. Undesired events are minimized through controlling the critical process parameters within the design space. Furthermore, preventative repair and maintenance actions are performed, such as the inspection and calibration of equipment. Published examples of this Quality by Design (QbD) approach include the mixing unit operation (Adam, Suzzi, Radeke, & Khinast, 2011) and the spray drying process (Baldinger, Clerdent, Rantanen, Yang, & Grohganz, 2012).

Mechanistic simulations are however rarely used for the risk assessment in the biotechnological industry (Rantanen & Khinast, 2015), even though PAT applications that consider various uncertainties provide probabilistic model predictions. They can therefore support risk-based decisions in real time (Stocker, Toschkoff, Sacher, & Khinast, 2014). A probabilistic bioprocess model that is applied in a soft sensor for monitoring and control of a production process predicts the probability distribution of the CQAs, and enables therefore the operators to react on a risk-based basis. For example, the cumulative distribution function that indicates the probability of not achieving the target (e.g. target yield) can be derived from the Monte Carlo simulation outputs (Fig. 3, Step 4: Risk quantification). The risk is then quantified as the product of the probability of the undesired event times the consequence of the undesired event (Cameron & Raman, 2005). The risk could be e.g. a loss of 0.5 kg product per cultivation or in economic perspectives e.g. $ 0.2 million. This then contributes to risk-based monitoring and controlling, and thereby improving both the performance and the safety of the production process (Li, Arellano-Garcia, & Wozny, 2008). The probability of reaching the target yield (and other QCAs) is continuously updated considering the uncertainties in the process variables. The process conditions

can then be controlled accordingly. In addition, the risk of failure of the cultivation process can be recognized early as it is updated in real-time based on the on-line measurements (H. Wang, Khan, Ahmed, & Imtiaz, 2016).

Both the Bayesian approach (Mockus, Laínez, Reklaitis, & Kirsch, 2011; Peterson & Lief, 2010) and the Monte Carlo approach (García-Muñoz, Luciani, Vaidyaraman, & Seibert, 2015) have been investigated and recommended for probabilistic PAT applications (Sin, Gernaey, & Lantz, 2009; Tabora & Domagalski, 2017). Depending on the purpose of the application, stochastic uncertainties might be incorporated in the model application. The probability of a specific event occurring, e.g. pump failure, could then be integrated (Barua, Gao, Pasman, & Mannan, 2016), hence engineers could assess the control algorithm robustness, and evaluate for example the response of the controller in case of a pump failure (Sin, Gernaey, & Lantz, 2009). Konakovsky et al. (2017) optimized the glucose feeding strategy for Chinese Hamster Ovary (CHO) cells in fed-batch culture, and considered various uncertainties in a Monte Carlo simulation assuming a worst-case scenario in silico. However, there is a considerable lack of published risk-based control implementations at large scale cultivations, especially using mechanistic models, despite their potential. One major challenge might be the time that is needed to develop and implement the model based control system. A second challenge might be regulatory validation requirements (Djuris & Djuric, 2017).

Heterogeneities at large-scale cultivations

Mechanistic models are applied to various PAT systems but homogeneous culture conditions are assumed usually, even though heterogeneities occur at large scale. In industrial large-scale bioprocesses, culture parameters like the substrate concentration, oxygen concentration, pH, and temperature are not homogeneously distributed, and gradients of them exist instead within the bioreactor (Lara et al., 2006). This is due to a lower mixing capability of large-scale bioreactors, which are in the scale of 10 – >300 m^3, compared to small laboratory-scale bioreactors (in the liter range). Cells circulating within large-scale bioreactors are consequently exposed to continuously changing conditions, and they need to adapt to these conditions constantly, which affects their metabolic activity. As a result, biomass and product yields are often lower at large scale than at lab scale (Bylund, Collet, Enfors, & Larsson, 1998; Zou, Xia, Chu, Zhuang, & Zhang, 2012). It is also possible that strains that showed a good productivity at lab scale do not grow at all at large scale. The challenge is therefore, to predict the behaviour of the microorganisms – mainly their productivity – at large-scale, without the need of expensive experiments at large scale to support optimal bioprocess development and control.

Gradients at large scale

Gradients occur at large scale because some components, like substrate, oxygen, and base or acid are added at one (or sometime several) positions locally to the cultivation, and it takes then several seconds up to minutes to distribute them homogeneously in the bioreactor (F. Delvigne, Destain, & Thonart, 2006). This leads to a heterogeneous distribution of culture parameters such as substrate, oxygen, carbon dioxide concentration, pH, and temperature (Fig. 4). Moreover, there exist pressure gradients in bioreactors due to the height of the bioreactor (Neubauer & Junne, 2016), and cells are exposed to changing flow conditions (Nienow, 2006, 2014). The effect of gradients in large-scale cultivations is often a reduced biomass yield and productivity compared to homogeneous lab-scale cultivations. Cells that travel through a large-scale bioreactor have to deal with changing conditions uninterruptedly, which means stress for the microorganisms. All microorganisms respond in a different way to the various

oscillating conditions. It is therefore necessary to study the strain specific response in so-called scale-down experiments, where the large-scale conditions are mimicked in lab-scale experiments so that the microbial response can be investigated (Neubauer & Junne, 2010; G. Wang et al., 2015). The effects of substrate, oxygen, and pH gradients have been studied in detail for many organisms and some examples will be highlighted in the following paragraphs. The objective of these studies is to investigate i) how the microorganisms respond to shifts in different conditions (or combinations of them); ii) how fast the organisms react to the changes on the transcriptomic, proteomic, and metabolomics level; and iii) what the long-term effects of oscillating conditions are.

Box 3 Mixing issues in large-scale bioreactors.

In large-scale bioreactors, there are usually heterogeneous conditions, e.g. substrate concentration gradients, because of lower mixing capabilities compared to the mixing in lab-scale reactors. For example, a substrate that is added locally to the bioreactor in a fed-batch cultivation is not distributed homogeneously in the bioreactor instantaneously (Fig. 4). Instead, a gradient is formed, i.e. there is a high substrate concentration close to the dosage point and a low concentration further away. This leads often to reduced product yields or titers as the cells that circulate through the bioreactor face constantly changing conditions. They need therefore to adapt to these changes, which requires energy that could otherwise be used to produce the product.

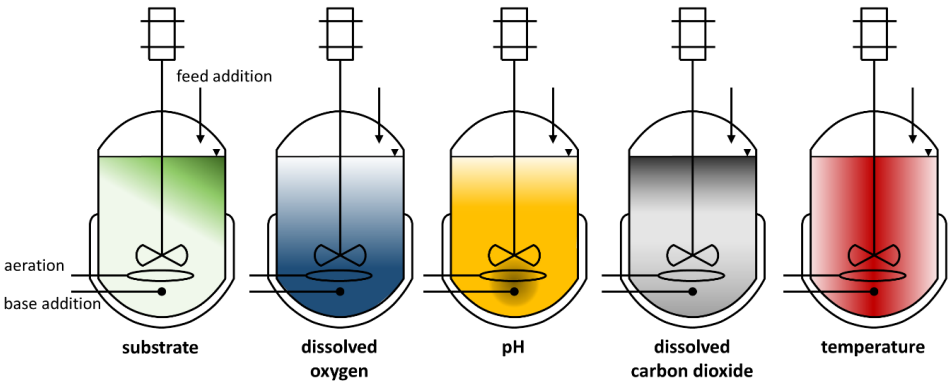


FIGURE 4 Gradients in large-scale cultivations. The formation of gradients for a fed-batch bioreactor configuration with the substrate feed addition at the top of the bioreactor, and the oxygen sparging and base addition at the bottom of the bioreactor. Dark color: high concentration/value; bright color: low concentration/value.

Substrate gradients play a major role in fed-batch cultivations, where the substrate is usually fed at the top of the bioreactor (Fig. 4, substrate). High substrate concentrations of up to 2 g L^{-1} have been reported near the feed addition point (Bylund et al., 1998). In the feeding zone, the cells take up a lot of substrate according to their maximum substrate uptake capacity. This leads to a fast depletion of oxygen there when high cell densities are reached, since the high substrate turnover demands a high respiratory capacity. A further consequence is the formation of glycolytic overflow metabolites, such as acetate, lactate, formate, and succinate in *E.coli* cultivations (Sunya, Bideaux, Molina-Jouve, & Gorret, 2013;

Xu, Jahic, Blomsten, & Enfors, 1999), or ethanol in *S. cerevisiae* cultivations (George, Larsson, & Enfors, 1993), as the entire carbon cannot be metabolized by the respiratory metabolism. However, these metabolites are then typically taken up again and re-metabolized in aerobic substrate-limited or -depleted zones, further away from the feed addition. Nevertheless, this generation and uptake of side products is a futile cycle, and can lead to a decreased biomass or product yield in the end, as these detours are less energetically efficient. Recent analyses investigated also the response on the transcriptomic and proteomic level of *E. coli* and *C. glutamicum* (Lemoine et al., 2016; Löffler et al., 2017; Simen et al., 2017; Sunya et al., 2013), which allow the detailed understanding of the microorganisms. This might allow to genetically engineer microorganisms that are less prone to stress, as shown in by Michalowski et al. (2017).

Oxygen gradients are particularly crucial when the cultivation reaches high cell densities and the available oxygen is consumed rapidly (Fig. 4, dissolved oxygen). Oxygen is added to the culture medium at the bottom of the bioreactor as gas (air or pure oxygen). Due to the low solubility of oxygen in water, the available oxygen is limited especially further away from the gas inlet, and in the feeding zone, where the substrate and oxygen uptake rates are high. Oscillating oxygen conditions have shown reduced recombinant protein production in *E.coli* (Sandoval-Basurto, Gosset, Bolívar, & Ramírez, 2005) and expression of non-proteinogenic branched-chain amino acid (Soini et al., 2008), which could lead to a misincorporation (an erroneous incorporation) into recombinant proteins. The lipase producing strain *Yarrowia lipolityca*, for example, shows a reduction of the lipase gene expression (Kar, Delvigne, Masson, Destain, & Thonart, 2008). Other strains, such as *Corynebacterium glutamicum*, seem to be robust with respect to substrate and oxygen gradients (Käß, Junne, Neubauer, Wiechert, & Oldiges, 2014; Limberg et al., 2017).

pH gradients occur in bioreactors because the pH of the cultivation broth is controlled by adding base or acid as needed by a pH control setup (Fig. 4, pH). In aerated cultivations, the acid/base is often added together with the incoming aeration gas that ensures a very fast distribution of the acid/base. In non-aerated cultivations, the acid/base is usually added to the bioreactor as liquid, either at the top or at another position, e.g. close to the impeller blades. The acid/base addition leads to a zone of unfavorable pH conditions in the close vicinity of the dosage point. Cells that pass this zone are prone to cell damage. pH shifts of almost 1 pH unit have been measured close to the alkali addition point in an 8 m^3 reactor for mammalian cell cultures (Langheinrich & Nienow, 1999), but even higher shifts might be expected. It was shown for *E.coli*, *B. subtilis*, and mammalian cells that oscillating pH conditions can have an effect on biomass growth, the metabolome, the transcriptome, and the cell viability (Amanullah, McFarlane, Emery, & Nienow, 2001; Brunner et al., 2017; Cortés, Flores, Bolívar, Lara, & Ramírez, 2016; Osman, Birch, & Varley, 2002). The extracellular pH has a direct impact on the cell physiology because it affects the intracellular pH, which is crucial for the enzymatic activity and controlled by proton pumps (Hansen et al., 2016).

CO_2 gradients have probably a major effect on the productivity of large-scale cultivations, too, but have been rarely investigated so far (Fig. 4, dissolved carbon dioxide) (Baez, Flores, Bolívar, & Ramírez, 2011; Blombach, Buchholz, Busche, Kalinowski, & Takors, 2013; Buchholz et al., 2014).

The effects of heterogeneities on the microbial activity and productivity are a combination of the fluid dynamics in the bioreactor and the microbial metabolism. Computational tools that describe the fluid dynamics are therefore coupled with bio-kinetic modelling in order to support the investigation of the effect of gradients at large scale.

Computational fluid dynamics in biotechnological cultivations

CFD simulations can predict among other things the motion of the fluid in bioreactors and are therefore applied e.g. for the prediction of the fluid dynamics and the mixing time when different impeller designs are tested for a bioreactor (Yang et al., 2012; Zou et al., 2012). Furthermore, CFD simulations can be combined with biokinetic models in order to investigate the effects attributed to gradients.

The fluid dynamics are numerically solved for the liquid volume in the CFD simulation. The fundamental mathematical equations are based on the conservation of mass, momentum, and energy. The most commonly used mathematical formulation is that of the Navier-Stokes equations. Further equations describing other phenomena, such as turbulence or eddy formation, are applied and solved depending on the aim of the simulation as well. The liquid volume is divided into many (up to a few million) small elements for the simulation, and the fluid dynamic equations are solved for each element.

CFD is applied for biotechnological cultivations because many key issues in cultivations are depending on the flow: mass transfer (e.g. mixing of feed streams; gas-liquid transfer), shear rates, and transport of the microorganisms through the bioreactor (Frank Delvigne, Takors, Mudde, van Gulik, & Noorman, 2017; Schmalzriedt, Jenne, Mauch, & Reuss, 2003). To this end, the biological kinetic equations are solved together with the fluid flow within the CFD simulations. The simulation then describes concentration gradients (e.g. substrate, dissolved oxygen), gradients of physical parameters (e.g. pH, mass transfer coefficients, gas hold-up), and the temporal and spatial performance of the microorganisms (e.g. substrate uptake, product formation, by-product formation, and growth etc.). The purpose of these studies is usually to predict the oscillating culture conditions at large scale, and to design controlled scale-down experiments at lab-scale that mimic large-scale conditions as closely as possible. Future interest are the prediction of the culture performance at large scale using CFD simulations.

As a first step, the fluid dynamics of the bioreactor, i.e. the velocity profile of the bioreactor, is simulated. Second, the mixing time is often simulated. The mixing time is the time that a tracer that was pulsed into the bioreactor needs to reach a homogeneous distribution (i.e. to achieve complete mixing) in the bioreactor. The predicted mixing time is used to validate the CFD simulation, since the simulated and experimental mixing time can be compared. Once the CFD simulation is validated, it can be applied to simulate microbial cultivations. The CFD simulation is then coupled with further gas-liquid transfer processes and biokinetic models to study bioprocess yields and performance in bioreactors.

There are two main approaches to model the microorganisms in the CFD simulation, which are applied depending on the purpose of the simulation: the Euler-Euler and the Euler-Lagrange approach. In the Euler-Euler approach, the microorganisms are treated as a continuum, i.e. all cells are treated in the same way as concentrations of a dissolved component (Bannari, Bannari, Vermette, & Proulx, 2012; Elqotbi, Vlaev, Montastruc, & Nikov, 2013). In the Euler-Lagrange approach, the fluid is treated as continuum, but the microorganisms are seen from the microbial point of view, i.e. individual cells travelling through the bioreactor are tracked (Haringa et al., 2016a; Lapin, Müller, & Reuss, 2004; Lapin, Schmid, & Reuss, 2006; Morchain, Gabelle, & Cockx, 2013). The advantage of the Euler-Euler approach is that it is computationally less demanding than the Euler-Lagrange approach. Nevertheless, this leads to a loss of the individual history of the cells (Lapin et al., 2004). This could be of interest when one wants to investigate, e.g. the effect of subsequent, unfavorable conditions or the culture conditions experienced by the microorganisms at large scale over time, so called lifelines of the cells (Haringa et al., 2016b; Kuschel, Siebler, & Takors, 2017). Scale-down experiments that resemble the large-scale conditions could be designed based on the CFD simulation. Haringa et al. (2017) recently proposed a 3-compartment lab-scale setup to mimic the conditions of a 22 m^3 *S. cerevisiae* cultivation,

which needs to be validated as a next step. CFD-based scale-down experiments will improve the quality of scale-down studies. CFD models are furthermore applied for example to enhance the treatment efficiency of wastewater plants (Karpinska & Bridgeman, 2016; Samstag et al., 2016).

Even though the CFD simulation of the fluid dynamics together with the biokinetic models is a powerful tool that provides accurate predictions, its application is limited to off-line studies. This is because the CFD simulation with its many mesh elements is computationally too demanding, especially when large-scale bioreactors need to be modelled. For on-line applications, such as monitoring and control, CFD simulations cannot be executed fast enough (e.g. every couple of minutes), and therefore compartment models are preferred.

Compartment models

Compartment models (CMs, also known as block models or network of zones) divide the liquid domain into a limited number of coarse elements, which is much lower than the number of elements used in the CFD simulation, and demands therefore less computational time compared to the CFD model. The properties (e.g. temperature, pH, concentrations) are assumed homogeneous in each compartment, and the Navier-Stokes equation is not included. Additional models, such as chemical, physical, biokinetic, or population balance models are solved only at an affordable simulation time. The compartments are interconnected with defined flow rates and the interface area between each adjacent compartment.

A consistent framework to build CMs is based on an initial CFD simulation that provides a validated steady state solution of the fluid dynamics (Le Moullec, Gentric, Potier, & Leclerc, 2010) (Fig. 5). The goal is to define the compartments based on the CFD results, preferable automatically according to a specific algorithm (Bezzo & Macchietto, 2004; Delafosse et al., 2010; Rigopoulos & Jones, 2003). Using the simulated velocity profile, compartments could be automatically defined, e.g. by identifying dead regions or recirculation regions (Bezzo & Macchietto, 2004). However, this is not a general solution, since adaptations for different kind of bioreactors (e.g. plug-flow reactors) or impeller configurations are needed. Moreover, the derived compartments do not necessarily represent for example a feeding or acid addition zone, which might be crucial for several applications. Bezzo & Macchietto (2004) have therefore proposed a hybrid algorithm that takes also other properties, e.g. concentrations, into account for the identification of compartments. This methodology has the disadvantage that also a CFD simulation including the bio-kinetics has to be performed once, in order to design the compartments. Nevertheless, the developed CFD-based CM can be validated in this way to replace the CFD simulation for accurate on-line applications. Manual definitions of compartments are also applied (Bashiri, Bertrand, & Chaouki, 2016; Nauha, Kálal, Ali, & Alopaeus, 2017). Delafosse et al. (2014) implemented homogeneously distributed compartments based on their coordinates but they needed more than 12000 compartments in order to match the CFD predicted and experimental mixing time of a 16 L bioreactor, which used more than 10^6 mesh elements for the CFD model. Such approaches will not be feasible for larger bioreactors or on-line applications combined with Monte-Carlo simulations (see Probabilistic model predictions).

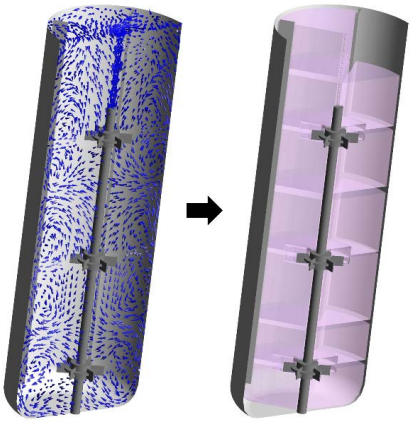


FIGURE 5 CFD-based compartment model. The steady-state velocity profile (left) can be used to define the compartments (right) for a compartment simulation. The velocity direction and magnitude (left, blue arrows), and an example of compartments (right, pink boxes) are shown for a bioreactor with 3 Rushton turbines and 4 baffles.

Once the compartments are defined, the volume flows (liquid and gaseous) between the compartments are taken from the CFD simulation. Further properties, such as volumes, pressure, shear rates, viscosities, etc. are also derived from the initial CFD simulation and used in the CM. Then, the CM needs to be validated, e.g. with mixing experiments or the already validated CFD simulation (Delafosse et al., 2015). Finally, biological, physical, and chemical kinetics can be incorporated, and then solved in all compartments together with the transfer processes between the compartments.

CMs have been applied to characterize the culture conditions in bioreactors (Bezzo, Macchietto, & Pantelides, 2003; Guha, Dudukovic, Ramachandran, Mehta, & Alvare, 2006; Le Moullec et al., 2010; Nauha et al., 2017; Rigopoulos & Jones, 2003; Wells & Ray, 2005; W. Zhao, Buffo, Alopaeus, Han, & Louhi-Kultanen, 2017). Furthermore, entire bioprocesses have been simulated. Glucose and acetate gradients could be predicted in a large-scale *E. coli* cultivation, and resembled the measured tendencies (Vrábel et al., 2001). An antibiotic producing *Streptomyces fradiae* cultivation in a 3 m^3 bioreactor was modelled and could be validated with dissolved oxygen measurements (Vlaev et al., 2000). The same framework has been applied to predict the dissolved oxygen tension in a 31 m^3 stirred tank bioreactor for the production of another antibiotic, and for a 236 m^3 bubble column for citric acid production, however without validation (Zahradník et al., 2001). Different gradients (substrate, oxygen, etc.) were predicted for wastewater treatment bioreactors (Alvarado, Vedantam, Goethals, & Nopens, 2012; Le Moullec et al., 2010; Rehman et al., 2017). These studies included the biokinetics of the ASM1 model (Henze, Gujer, Mino, & Lossdreacht, 2000) to describe the microbial conversions in the system in very large aeration tanks of several 1000 m^3. For a 30 m^3 *E. coli* cultivation gradients for the glucose concentration, the growth rate, and others were predicted, whereas a population balance was applied to account for cell heterogeneities (Pigou & Morchain, 2015; Pigou, Morchain, Fede, Penet, & Laronze, 2017).

Conclusions

Several PAT applications for cultivation processes have reached a mature level. Models have been increasingly applied to simulate biotechnological production processes, such as large-scale bio-pharmaceutical antibiotic manufacturing processes. The models are applied for different off-line and on-line purposes including process development, optimization, monitoring, and control. This contribution has focused on mechanistic models because of their advantages over data-driven models. As mechanistic models have e.g. larger extrapolation capabilities, they can predict scenarios often accurately where there is little or no data available. This is of special interest if new process conditions should be tested without performing costly experiments.

Uncertainties are considered in the model predictions for risk assessment, e.g. by performing Monte Carlo simulations. Various uncertainties have been considered: model input uncertainties, which include uncertain model parameters and variable medium concentrations, and stochastic uncertainties such as pump failures. Applications taking uncertainties into account provide probabilistic model predictions that allow risk-based decision making. However, uncertainties have been included only in one-compartment simulations yet.

Heterogeneous process conditions, such as substrate gradients in fed-batch cultivations have been successfully studied and predicted with CFD simulations. To this end, bio-kinetic models have been integrated in the CFD simulations to track spatial and temporal changes of process conditions. Appropriately scaled-down bioreactors that mimic large-scale conditions are designed based on CFD simulations. This consistent approach enables researchers to test the effects of large-scale conditions in lab-scale bioreactors at the early stage of strain screening and process development. Interactions of the microorganisms and process conditions could be understood in more detail. Since CFD simulations are computationally demanding, compartment models that reduce the spatial resolution and thereby the computational calculation demand have been developed, and are promising for on-line applications.

Perspectives

So far, CFD and compartment models that were coupled with bio-kinetic models have been applied for off-line applications including optimization of process design and scale-up as well as increasing the process understanding. However, heterogeneities have not been considered for on-line applications, such as process monitoring and control. It is important to consider the heterogeneous process conditions that occur at large scale because traditional one-point measurements, e.g. of dissolved oxygen, might not be optimal for on-line control. New on-line control algorithms could consider the predicted gradients in bioreactors, and hence provide an optimized control for the bioprocess as well as propose optimized location for sensors to collect online data. Furthermore, uncertainties need to be considered for risk assessment, e.g. by Monte Carlo simulations, to allow a risk-based decision making. The latter is particularly useful for operating microbial cultivations in an optimal sense so that operators can take early and timely decisions in case process deviations are predicted to lead to sub-optimal process yields or outright failures of a batch/fed-batch/continuous process. Simulations considering both heterogeneities and uncertainties are therefore needed for risk-based on-line control applications to achieve optimized large-scale bioprocess operations. For these on-line applications, the model predictions are required in a timeframe in the order of minutes to allow control actions in real-time. CFD-based compartment models can therefore be a good alternative over CFD models for on-line applications, as they can be solved in real time.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project).

References

Adam, S., Suzzi, D., Radeke, C., & Khinast, J. G. (2011). An integrated Quality by Design (QbD) approach towards design space definition of a blending unit operation by Discrete Element Method (DEM) simulation. *European Journal of Pharmaceutical Sciences*, *42*(1–2), 106–115. http://doi.org/10.1016/j.ejps.2010.10.013

Albaek, M. O., Gernaey, K. V., Hansen, M. S., & Stocks, S. M. (2011). Modeling enzyme production with Aspergillus oryzae in pilot scale vessels with different agitation, aeration, and agitator types. *Biotechnology and Bioengineering*, *108*(8), 1828–1840. http://doi.org/10.1002/bit.23121

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology - Improving cell factory performance. *Metabolic Engineering*, *24*, 38–60. http://doi.org/10.1016/j.ymben.2014.03.007

Alvarado, A., Vedantam, S., Goethals, P., & Nopens, I. (2012). A compartmental model to describe hydraulics in a full-scale waste stabilization pond. *Water Research*, *46*(2), 521–530. http://doi.org/10.1016/j.watres.2011.11.038

Alves-Rausch, J., Bienert, R., Grimm, C., & Bergmaier, D. (2014). Real time in-line monitoring of large scale Bacillus fermentations with near-infrared spectroscopy. *Journal of Biotechnology*, *189*, 120–128. http://doi.org/10.1016/j.jbiotec.2014.09.004

Amanullah, A., McFarlane, C. M., Emery, A. N., & Nienow, A. W. (2001). Scale-down model to simulate spatial pH variations in large-scale bioreactors. *Biotechnology and Bioengineering*, *73*(5), 390–399. http://doi.org/10.1002/bit.1072

Baez, A., Flores, N., Bolívar, F., & Ramírez, O. T. (2011). Simulation of dissolved CO2 gradients in a scale-down system: A metabolic and transcriptional study of recombinant Escherichia coli. *Biotechnology Journal*, *6*(8), 959–967. http://doi.org/10.1002/biot.201000407

Baldinger, A., Clerdent, L., Rantanen, J., Yang, M., & Grohganz, H. (2012). Quality by design approach in the optimization of the spray-drying process. *Pharmaceutical Development and Technology*, *17*(4), 389–397. http://doi.org/10.3109/10837450.2010.550623

Bannari, R., Bannari, A., Vermette, P., & Proulx, P. (2012). A model for cellulase production from Trichoderma reesei in an airlift reactor. *Biotechnology and Bioengineering*, *109*(8), 2025–2038. http://doi.org/10.1002/bit.24473

Barua, S., Gao, X., Pasman, H., & Mannan, M. S. (2016). Bayesian network based dynamic operational risk assessment. *Journal of Loss Prevention in the Process Industries*, *41*, 399–410. http://doi.org/10.1016/j.jlp.2015.11.024

Bashiri, H., Bertrand, F., & Chaouki, J. (2016). Development of a multiscale model for the design and scale-up of gas/liquid stirred tank reactors. *Chemical Engineering Journal*, *297*, 277–294. http://doi.org/10.1016/j.cej.2016.03.102

Belini, V. L., Wiedemann, P., & Suhr, H. (2013). In situ microscopy: A perspective for industrial bioethanol production monitoring. *Journal of Microbiological Methods*, *93*(3), 224–232. http://doi.org/10.1016/j.mimet.2013.03.009

Bezzo, F., & Macchietto, S. (2004). A general methodology for hybrid multizonal/CFD models: Part II. Automatic zoning. *Computers and Chemical Engineering*, *28*(4), 513–525. http://doi.org/10.1016/j.compchemeng.2003.08.010

Bezzo, F., Macchietto, S., & Pantelides, C. C. (2003). General hybrid multizonal/CFD approach for bioreactor modeling. *AIChE Journal*, *49*(8), 2133–2148. http://doi.org/10.1002/aic.690490821

Biechele, P., Busse, C., Solle, D., Scheper, T., & Reardon, K. (2015). Sensor systems for bioprocess monitoring. *Engineering in Life Sciences*, *15*(5), 469–488. http://doi.org/10.1002/elsc.201500014

Blombach, B., Buchholz, J., Busche, T., Kalinowski, J., & Takors, R. (2013). Impact of different CO2/HCO3- levels on metabolism and regulation in Corynebacterium glutamicum. *Journal of Biotechnology*, *168*(4), 331–340. http://doi.org/10.1016/j.jbiotec.2013.10.005

Bluma, A., Höpfner, T., Lindner, P., Rehbock, C., Beutel, S., Riechers, D., … Scheper, T. (2010). In-situ imaging sensors for bioprocess monitoring: State of the art. *Analytical and Bioanalytical Chemistry*, *398*(6), 2429–2438. http://doi.org/10.1007/s00216-010-4181-y

Brunner, M., Braun, P., Doppler, P., Posch, C., Behrens, D., Herwig, C., & Fricke, J. (2017). The impact of pH inhomogeneities on CHO cell physiology and fed-batch process performance – two-compartment scale-down modelling and intracellular pH excursion. *Biotechnology Journal*, *12*(7). http://doi.org/10.1002/biot.201600633

Buchholz, J., Graf, M., Freund, A., Busche, T., Kalinowski, J., Blombach, B., & Takors, R. (2014). CO2/HCO3- perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in Corynebacterium glutamicum. *Applied Microbiology and Biotechnology*, *98*(20), 8563–8572. http://doi.org/10.1007/s00253-014-6014-y

Bylund, F., Collet, E., Enfors, S.-O., & Larsson, G. (1998). Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Engineering*, *18*(3), 171. http://doi.org/10.1007/s004490050427

Cameron, I., & Raman, R. (2005). *Process Systems Risk Management*. Elsevier Academic Press.

Campbell, K., Xia, J., & Nielsen, J. (2017). The Impact of Systems Biology on Bioprocessing. *Trends in Biotechnology*, *xx*, 1–13. http://doi.org/10.1016/j.tibtech.2017.08.011

Chis, O. T., Banga, J. R., & Balsa-Canto, E. (2011). Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE*, *6*(11). http://doi.org/10.1371/journal.pone.0027755

Chisti, Y., & Moo-Young, M. (2003). Bioreactors. In *Encyclopedia of Physical Science and Technology* (pp. 247–271). Elsevier. http://doi.org/10.1016/B0-12-227410-5/00067-3

Cortés, J. T., Flores, N., Bolívar, F., Lara, A. R., & Ramírez, O. T. (2016). Physiological effects of pH gradients on Escherichia coli during plasmid DNA production. *Biotechnology and Bioengineering*, *113*(3), 598–611. http://doi.org/10.1002/bit.25817

Craven, S., Whelan, J., & Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *Journal of Process Control*, *24*(4), 344–357. http://doi.org/10.1016/j.jprocont.2014.02.007

Delafosse, A., Calvo, S., Collignon, M. L., Delvigne, F., Crine, M., & Toye, D. (2015). Euler-Lagrange approach to model heterogeneities in stirred tank bioreactors - Comparison to experimental flow characterization and particle tracking. *Chemical Engineering Science*, *134*, 457–466. http://doi.org/10.1016/j.ces.2015.05.045

Delafosse, A., Collignon, M. L., Calvo, S., Delvigne, F., Crine, M., Thonart, P., & Toye, D. (2014). CFD-based compartment model for description of mixing in bioreactors. *Chemical Engineering*

Science, *106*, 76–85. http://doi.org/10.1016/j.ces.2013.11.033

Delafosse, A., Delvigne, F., Collignon, M. L., Crine, M., Thonart, P., & Toye, D. (2010). Development of a compartment model based on CFD simulations for description of mixing in bioreactors. *Biotechnology, Agronomy and Society and Environment*, *14*(SPEC. ISSUE 2), 517–522.

Delvigne, F., Destain, J., & Thonart, P. (2006). A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. *Biochemical Engineering Journal*, *28*(3), 256–268. http://doi.org/10.1016/j.bej.2005.11.009

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., & Noorman, H. (2017). Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. *Microbial Biotechnology*, *10*(5), 1267–1274. http://doi.org/10.1111/1751-7915.12803

Djuris, J., & Djuric, Z. (2017). Modeling in the quality by design environment: Regulatory requirements and recommendations for design space and control strategy appointment. *International Journal of Pharmaceutics*, *533*(2), 346–356. http://doi.org/10.1016/j.ijpharm.2017.05.070

Elqotbi, M., Vlaev, S. D., Montastruc, L., & Nikov, I. (2013). CFD modelling of two-phase stirred bioreaction systems by segregated solution of the Euler-Euler model. *Computers and Chemical Engineering*, *48*, 113–120. http://doi.org/10.1016/j.compchemeng.2012.08.005

Esener, A. A., Roels, J. A., & Kossen, N. W. F. (1983). Theory and applications of unstructured growth models: Kinetic and energetic aspects. *Biotechnology and Bioengineering*, *25*(12), 2803–2841. http://doi.org/10.1002/bit.260251202

Fayolle, P., Picque, D., & Corrieu, G. (1997). Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy. *Vibrational Spectroscopy*, *14*(2), 247–252. http://doi.org/10.1016/s0924-2031(97)00004-0

FDA. (2004). *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. U.S. Food and Drug Administration, U.S. Department of Health and Human Services.

Ferrari, A., Gutiérrez, S., & Sin, G. (2016). Modeling a production scale milk drying process: Parameter estimation, uncertainty and sensitivity analysis. *Chemical Engineering Science*, *152*, 301–310. http://doi.org/10.1016/j.ces.2016.06.019

Flores-Alsina, X., Solon, K., Kazadi Mbamba, C., Tait, S., Gernaey, K. V., Jeppsson, U., & Batstone, D. J. (2016). Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes. *Water Research*, *95*(1), 370–382. http://doi.org/10.1016/j.watres.2016.03.012

Frahm, B. (2014). Seed Train Optimization for Cell Culture (pp. 355–367). http://doi.org/10.1007/978-1-62703-733-4_22

García-Muñoz, S., Luciani, C. V., Vaidyaraman, S., & Seibert, K. D. (2015). Definition of Design Spaces Using Mechanistic Models and Geometric Projections of Probability Maps. *Organic Process Research and Development*, *19*(8), 1012–1023. http://doi.org/10.1021/acs.oprd.5b00158

George, S., Larsson, G., & Enfors, S. O. (1993). A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of Saccharomyces cerevisiae. *Bioprocess Engineering*, *9*(6), 249–257. http://doi.org/10.1007/BF01061530

Gernaey, K. V., Cervera-Padrell, A. E., & Woodley, J. M. (2012). A perspective on PSE in pharmaceutical process development and innovation. *Computers and Chemical Engineering*, *42*, 15–29. http://doi.org/10.1016/j.compchemeng.2012.02.022

Gernaey, K. V, Lantz, A. E., Tufvesson, P., Woodley, J. M., & Sin, G. (2010). Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in Biotechnology*, *28*(7), 346–54. http://doi.org/10.1016/j.tibtech.2010.03.006

Glassey, J., Gernaey, K. V., Clemens, C., Schulz, T. W., Oliveira, R., Striedner, G., & Mandenius, C. F. (2011). Process analytical technology (PAT) for biopharmaceuticals. *Biotechnology Journal*, *6*(4), 369–377. http://doi.org/10.1002/biot.201000356

Golabgir, A., & Herwig, C. (2016). Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production. *Chemie-Ingenieur-Technik*, *88*(6), 764–776. http://doi.org/10.1002/cite.201500101

Goldrick, S., Stefan, A., Lovett, D., Montague, G., & Lennox, B. (2015). The development of an industrial-scale fed-batch fermentation simulation. *Journal of Biotechnology*, *193*, 70–82. http://doi.org/10.1016/j.jbiotec.2014.10.029

Guha, D., Dudukovic, M. P., Ramachandran, P. A., Mehta, S., & Alvare, J. (2006). CFD-based compartmental modeling of single phase stirred-tank reactors. *AIChE Journal*, *52*(5), 1836–1846. http://doi.org/10.1002/aic.10772

Halton, J. H. (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, *7*(12), 701–702. http://doi.org/10.1145/355588.365104

Hansen, G., Johansen, C. L., Marten, G., Wilmes, J., Jespersen, L., & Arneborg, N. (2016). Influence of extracellular pH on growth, viability, cell size, acidification activity, and intracellular pH of Lactococcus lactis in batch fermentations. *Applied Microbiology and Biotechnology*, *100*(13), 5965–5976. http://doi.org/10.1007/s00253-016-7454-3

Haringa, C., Deshmukh, A. T., Mudde, R. F., & Noorman, H. J. (2017). Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chemical Engineering Science*, *170*, 653–669. http://doi.org/10.1016/j.ces.2017.01.014

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., … Noorman, H. J. (2016a). Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Engineering in Life Sciences*, *16*(7), 652–663. http://doi.org/10.1002/elsc.201600061

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., … Noorman, H. J. (2016b). Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Engineering in Life Sciences*, *16*(7), 652–663. http://doi.org/10.1002/elsc.201600061

Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, *81*(1), 23–69. http://doi.org/10.1016/S0951-8320(03)00058-9

Henze, M., Gujer, W., Mino, T., & Lossdreacht, M. (2000). *Activated sludge models ASM1, ASM2, ASM2d and ASM3*. IWA Publishing Company.

ICH Q8(R2). (2009). Pharmaceutical Development. Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf

ICH Q9. (2005). Quality Risk Management. Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf

Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among

input variables. *Communications in Statistics - Simulation and Computation*, *11*(3), 311–334. http://doi.org/10.1080/03610918208812265

Jiménez-Hornero, J. E., Santos-Dueñas, I. M., & García-García, I. (2009). Optimization of biotechnological processes. The acetic acid fermentation. Part III: Dynamic optimization. *Biochemical Engineering Journal*, *45*(1), 22–29. http://doi.org/10.1016/j.bej.2009.01.011

Johnsson, O. (2015). *Perturbation-based Control of Industrial Fed-batch Bioprocesses*. Lund University.

Junker, B. (2004). Fermentation. In *Kirk-Othmer Encyclopedia of Chemical Technology*. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/0471238961.0605181319051407.a01.pub2

Kadlec, P., Gabrys, B., & Strandt, S. (2009). Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering*, *33*(4), 795–814. http://doi.org/10.1016/j.compchemeng.2008.12.012

Kar, T., Delvigne, F., Masson, M., Destain, J., & Thonart, P. (2008). Investigation of the effect of different extracellular factors on the lipase production by Yarrowia lipolityca on the basis of a scale-down approach. *Journal of Industrial Microbiology and Biotechnology*, *35*(9), 1053–1059. http://doi.org/10.1007/s10295-008-0382-1

Karpinska, A. M., & Bridgeman, J. (2016). CFD-aided modelling of activated sludge systems - A critical review. *Water Research*, *88*, 861–879. http://doi.org/10.1016/j.watres.2015.11.008

Käß, F., Junne, S., Neubauer, P., Wiechert, W., & Oldiges, M. (2014). Process inhomogeneity leads to rapid side product turnover in cultivation of Corynebacterium glutamicum. *Microbial Cell Factories*, *13*(1). http://doi.org/10.1186/1475-2859-13-6

Konakovsky, V., Clemens, C., Müller, M. M., Bechmann, J., & Herwig, C. (2017). A robust feeding strategy to maintain set-point glucose in mammalian fed-batch cultures when input parameters have a large error. *Biotechnology Progress*, *33*(2), 317–336. http://doi.org/10.1002/btpr.2438

Koutinas, M., Kiparissides, A., Pistikopoulos, E. N., & Mantalaris, A. (2012). Bioprocess Systems Engineering: Transferring Traditional Process Engineering Principles To Industrial Biotechnology. *Computational and Structural Biotechnology Journal*, *3*(4), e201210022. http://doi.org/10.5936/csbj.201210022

Kroll, P., Hofer, A., Stelzer, I. V., & Herwig, C. (2017). Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. *Process Biochemistry*, *62*(July), 24–36. http://doi.org/10.1016/j.procbio.2017.07.017

Kuschel, M., Siebler, F., & Takors, R. (2017). Lagrangian Trajectories to Predict the Formation of Population Heterogeneity in Large-Scale Bioreactors. *Bioengineering*, *4*(2), 27. http://doi.org/10.3390/bioengineering4020027

Langheinrich, C., & Nienow, A. W. (1999). Control of pH in large-scale, free suspension animal cell bioreactors: Alkali addition and pH excursions. *Biotechnology and Bioengineering*, *66*(3), 171–179. http://doi.org/10.1002/(SICI)1097-0290(1999)66:3<171::AID-BIT5>3.0.CO;2-T

Lapin, A., Müller, D., & Reuss, M. (2004). Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler−Lagrange Methods: Traveling along the Lifelines of Single Cells †. *Industrial & Engineering Chemistry Research*, *43*(16), 4647–4656. http://doi.org/10.1021/ie030786k

Lapin, A., Schmid, J., & Reuss, M. (2006). Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor-A structured-segregated approach. *Chemical Engineering Science*, *61*(14), 4783–4797. http://doi.org/10.1016/j.ces.2006.03.003

Lara, A. R., Galindo, E., Ramírez, O. T., & Palomares, L. a. (2006). Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Molecular Biotechnology*, *34*(3), 355–381. http://doi.org/10.1385/MB:34:3:355

Le Moullec, Y., Gentric, C., Potier, O., & Leclerc, J. P. (2010). Comparison of systemic, compartmental and CFD modelling approaches: Application to the simulation of a biological reactor of wastewater treatment. *Chemical Engineering Science*, *65*(1), 343–350. http://doi.org/10.1016/j.ces.2009.06.035

Lemoine, A., Limberg, M. H., Kästner, S., Oldiges, M., Neubauer, P., & Junne, S. (2016). Performance loss of *Corynebacterium glutamicum* cultivations under scale-down conditions using complex media. *Engineering in Life Sciences*, 1–13. http://doi.org/10.1002/elsc.201500144

Lencastre Fernandes, R., Bodla, V. K., Carlquist, M., Heins, A.-L., Eliasson Lantz, A., Sin, G., & Gernaey, K. V. (2013). Applying Mechanistic Models in Bioprocess Development. In *Advances in biochemical engineering/biotechnology* (Vol. 132, pp. 137–166). http://doi.org/10.1007/10_2012_166

Li, P., Arellano-Garcia, H., & Wozny, G. (2008). Chance constrained programming approach to process optimization under uncertainty. *Computers and Chemical Engineering*, *32*(1–2), 25–45. http://doi.org/10.1016/j.compchemeng.2007.05.009

Limberg, M. H., Schulte, J., Aryani, T., Mahr, R., Baumgart, M., Bott, M., … Oldiges, M. (2017). Metabolic profile of 1,5-diaminopentane producing Corynebacterium glutamicum under scale-down conditions: Blueprint for robustness to bioreactor inhomogeneities. *Biotechnology and Bioengineering*, *114*(3), 560–575. http://doi.org/10.1002/bit.26184

Loewenthal, R. E., Ekama, G. A., & Marais, G. R. (1989). Mixed weak acid/base systems Part I - Mixture characterisation. *Water SA*, *15*(1), 3–24.

Löffler, M., Simen, J. D., Müller, J., Jäger, G., Laghrami, S., Schäferhoff, K., … Takors, R. (2017). Switching between nitrogen and glucose limitation: Unraveling transcriptional dynamics in Escherichia coli. *Journal of Biotechnology*, *258*(January), 2–12. http://doi.org/10.1016/j.jbiotec.2017.04.011

Luttmann, R., Bracewell, D. G., Cornelissen, G., Gernaey, K. V., Glassey, J., Hass, V. C., … Mandenius, C. F. (2012). Soft sensors in bioprocessing: A status report and recommendations. *Biotechnology Journal*, *7*(8), 1040–1048. http://doi.org/10.1002/biot.201100506

Mandenius, C.-F. (Ed.). (2016). *Bioreactors: Design, Operation and Novel Applications*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. http://doi.org/10.1002/9783527683369

Mandenius, C. F., & Gustavsson, R. (2015). Mini-review: Soft sensors as means for PAT in the manufacture of bio-therapeutics. *Journal of Chemical Technology and Biotechnology*, *90*(2), 215–227. http://doi.org/10.1002/jctb.4477

Markopoulos, J., Christofi, C., & Katsinaris, I. (2007). Mass Transfer Coefficients in Mechanically Agitated Gas-Liquid Contactors. *Chemical Engineering & Technology*, *30*(7), 829–834. http://doi.org/10.1002/ceat.200600394

Marquard, D., Schneider-Barthold, C., Düsterloh, S., Scheper, T., & Lindner, P. (2017). Online monitoring of cell concentration in high cell density Escherichia coli cultivations using in situ Microscopy. *Journal of Biotechnology*, *259*(May), 83–85. http://doi.org/10.1016/j.jbiotec.2017.08.003

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2),

239–245. http://doi.org/10.1080/00401706.1979.10489755

McKay, M. D., Morrison, J. D., & Upton, S. C. (1999). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, *117*(1), 44–51. http://doi.org/10.1016/S0010-4655(98)00155-6

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., & Gernaey, K. V. (2017a). Application of a mechanistic model as a tool for on-line monitoring of pilot scale filamentous fungal fermentation processes—The importance of evaporation effects. *Biotechnology and Bioengineering*, *114*(3), 589–599. http://doi.org/10.1002/bit.26187

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., & Gernaey, K. V. (2017b). Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends in Biotechnology*, *35*(10), 914–924. http://doi.org/10.1016/j.tibtech.2017.07.002

Meng, X. (2013). Scalable simple random sampling and stratified sampling. *30th International Conference on Machine Learning, Icml 2013*, (2).

Mercier, S. M., Diepenbroek, B., Wijffels, R. H., & Streefland, M. (2014). Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations. *Trends in Biotechnology*, *32*(6), 329–336. http://doi.org/10.1016/j.tibtech.2014.03.008

Meyer, H.-P., Minas, W., & Schmidhalter, D. (2016). Industrial-Scale Fermentation. In *Industrial Biotechnology* (pp. 1–53). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. http://doi.org/10.1002/9783527807833.ch1

Michalowski, A., Siemann-Herzberg, M., & Takors, R. (2017). Escherichia coli HGT: Engineered for high glucose throughput even under slowly growing or resting conditions. *Metabolic Engineering*, *40*(December 2016), 93–103. http://doi.org/10.1016/j.ymben.2017.01.005

Mockus, L., Laínez, J. M., Reklaitis, G., & Kirsch, L. (2011). A Bayesian Approach to Pharmaceutical Product Quality Risk Quantification. *Informatica*, *22*(4), 537–558.

Monod, J. (1949). The Growth of Bacterial Cultures. *Annual Review of Microbiology*, *3*(1), 371–394. http://doi.org/10.1146/annurev.mi.03.100149.002103

Morchain, J., Gabelle, J.-C., & Cockx, A. (2013). Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. *AIChE Journal*, *59*(2), 369–379. http://doi.org/10.1002/aic.13820

Musvoto, E. V., Wentzel, M. C., Loewenthal, R. E., & Ekama, G. A. (2000). Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems. *Water Research*, *34*(6), 1857–1867. http://doi.org/10.1016/S0043-1354(99)00334-6

Nauha, E. K., Kálal, Z., Ali, J. M., & Alopaeus, V. (2017). Compartmental modeling of large stirred tank bioreactors with high gas volume fractions. *Chemical Engineering Journal*, *334*(November 2017), 2319–2334. http://doi.org/10.1016/j.cej.2017.11.182

Nelsen, R. B. (2006). *An introduction to copulas*. Springer.

Neubauer, P., Cruz, N., Glauche, F., Junne, S., Knepper, A., & Raven, M. (2013). Consistent development of bioprocesses from microliter cultures to the industrial scale. *Engineering in Life Sciences*, *13*(3), 224–238. http://doi.org/10.1002/elsc.201200021

Neubauer, P., & Junne, S. (2010). Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Current Opinion in Biotechnology*, *21*(1), 114–121. http://doi.org/10.1016/j.copbio.2010.02.001

Neubauer, P., & Junne, S. (2016). Scale-Up and Scale-Down Methodologies for Bioreactors. In C. F. Mandenius (Ed.), *Bioreactors: Design, Operation and Novel Applications* (pp. 323–354). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. http://doi.org/10.1002/9783527683369.ch11

Nielsen, J. (2003). Metabolic Engineering. In *Encyclopedia of Physical Science and Technology* (pp. 391–406). Elsevier. http://doi.org/10.1016/B0-12-227410-5/00422-1

Nienow, A. W. (2006). Reactor engineering in large scale animal cell culture. *Cytotechnology*, *50*(1–3), 9–33. http://doi.org/10.1007/s10616-006-9005-8

Nienow, A. W. (2014). Re "Development of a scale-down model of hydrodynamic stress to study the performance of an industrial CHO cell line under simulated production scale bioreactor conditions" [Sieck, J.B., Cordes, T., Budach, W.E., Rhiel, M.H., Suemeghy, Z., Leist, C., Vill. *Journal of Biotechnology*, *171*(1), 82–84. http://doi.org/10.1016/j.jbiotec.2013.12.002

Ödman, P., Johansen, C. L., Olsson, L., Gernaey, K. V., & Lantz, A. E. (2009). On-line estimation of biomass, glucose and ethanol in Saccharomyces cerevisiae cultivations using in-situ multi-wavelength fluorescence and software sensors. *Journal of Biotechnology*, *144*(2), 102–112. http://doi.org/10.1016/j.jbiotec.2009.08.018

Osman, J. J., Birch, J., & Varley, J. (2002). The response of GS-NS0 myeloma cells to single and multiple pH perturbations. *Biotechnology and Bioengineering*, *79*(4), 398–407. http://doi.org/10.1002/bit.10198

Pais, D. A. M., Carrondo, M. J. T., Alves, P. M., & Teixeira, A. P. (2014). Towards real-time monitoring of therapeutic protein quality in mammalian cell processes. *Current Opinion in Biotechnology*, *30*, 161–167. http://doi.org/10.1016/j.copbio.2014.06.019

Peterson, J. J., & Lief, K. (2010). The ICH Q8 Definition of Design Space: A Comparison of the Overlapping Means and the Bayesian Predictive Approaches. *Statistics in Biopharmaceutical Research*, *2*(2), 249–259. http://doi.org/10.1198/sbr.2009.08065

Pigou, M., & Morchain, J. (2015). Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. *Chemical Engineering Science*, *126*, 267–282. http://doi.org/10.1016/j.ces.2014.11.035

Pigou, M., Morchain, J., Fede, P., Penet, M. I., & Laronze, G. (2017). An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors. *Chemical Engineering Science*, *171*, 218–232. http://doi.org/10.1016/j.ces.2017.05.026

Posch, A. E., Herwig, C., & Spadiut, O. (2013). Science-based bioprocess design for filamentous fungi. *Trends in Biotechnology*, *31*(1), 37–44. http://doi.org/10.1016/j.tibtech.2012.10.008

Process Development Forum. (2014). Mechanistic Modeling: Does it Have a Future in Process Development? Retrieved December 12, 2017, from http://www.processdevelopmentforum.com/articles/mechanistic-modeling-does-it-have-a-future-in-process-development-2/

Prunescu, R. M., Blanke, M., Jakobsen, J. G., & Sin, G. (2015). Dynamic modeling and validation of a biomass hydrothermal pretreatment process-a demonstration scale study. *AIChE Journal*, (August), n/a-n/a. http://doi.org/10.1002/aic.14954

Rantanen, J., & Khinast, J. (2015). The Future of Pharmaceutical Manufacturing Sciences. *Journal of Pharmaceutical Sciences*, *104*(11), 3612–3638. http://doi.org/10.1002/jps.24594

Rathore, A. S., Mittal, S., Pathak, M., & Arora, A. (2014). Guidance for performing multivariate data analysis of bioprocessing data: Pitfalls and recommendations. *Biotechnology Progress*, *30*(4),

967–973. http://doi.org/10.1002/btpr.1922

Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. *Nature Biotechnology*, *27*(1), 26–34. http://doi.org/10.1038/nbt0109-26

Rehman, U., Audenaert, W., Amerlinck, Y., Maere, T., Arnaldos, M., & Nopens, I. (2017). How well-mixed is well mixed? Hydrodynamic-biokinetic model integration in an aerated tank of a full-scale water resource recovery facility. *Water Science and Technology*, *76*(8), 1950–1965. http://doi.org/10.2166/wst.2017.330

Rigopoulos, S., & Jones, A. (2003). A hybrid CFD-reaction engineering framework for multiphase reactor modelling: Basic concept and application to bubble column reactors. *Chemical Engineering Science*, *58*(14), 3077–3089. http://doi.org/10.1016/S0009-2509(03)00179-9

Roels, J. A. (1981). THE APPLICATION OF MACROSCOPIC PRINCIPLES TO MICROBIAL METABOLISM. *Annals of the New York Academy of Sciences*, *369*(1 Biochemical E), 113–134. http://doi.org/10.1111/j.1749-6632.1981.tb14182.x

Sagmeister, P., Kment, M., Wechselberger, P., Meitz, A., Langemann, T., & Herwig, C. (2013). Soft-sensor assisted dynamic investigation of mixed feed bioprocesses. *Process Biochemistry*, *48*(12), 1839–1847. http://doi.org/10.1016/j.procbio.2013.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley \& Sons, Ltd. http://doi.org/10.1002/9780470725184

Samstag, R. W., Ducoste, J. J., Griborio, A., Nopens, I., Batstone, D. J., Wicks, J. D., … Laurent, J. (2016). CFD for wastewater treatment: An overview. *Water Science and Technology*, *74*(3), 549–563. http://doi.org/10.2166/wst.2016.249

Sandoval-Basurto, E. A., Gosset, G., Bolívar, F., & Ramírez, O. T. (2005). Culture of Escherichia coli under dissolved oxygen gradients simulated in a two-compartment scale-down system: Metabolic response and production of recombinant protein. *Biotechnology and Bioengineering*, *89*(4), 453–463. http://doi.org/10.1002/bit.20383

Schmalzriedt, S., Jenne, M., Mauch, K., & Reuss, M. (2003). Integration of Physiology and Fluid Dynamics. In U. von Stockar, L. A. M. van der Wielen, A. Bruggink, J. M. S. Cabral, S.-O. Enfors, P. Fernandes, … A. J. J. Straathof (Eds.), *Process Integration in Biochemical Engineering* (pp. 19–68). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/3-540-36782-9_2

Simen, J. D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., … Broicher, A. (2017). Transcriptional response of Escherichia coli to ammonia and glucose fluctuations. *Microbial Biotechnology*, *10*(4), 858–872. http://doi.org/10.1111/1751-7915.12713

Sin, G., & Gernaey, K. V. (2016). Data Handling and Parameter Estimation. In M. C. M. van Loosdrecht, P. H. Nielsen, C. M. Lopez-Vazquez, & D. Brdjanovic (Eds.), *Experimental Methods in Wastewater Treatment* (pp. 201–234). IWA Publishing Company.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M., & Gujer, W. (2009). Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, *43*(11), 2894–2906. http://doi.org/10.1016/j.watres.2009.03.048

Sin, G., Gernaey, K. V, & Lantz, A. E. (2009). Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. *Biotechnology Progress*, *25*(4), 1043–1053. http://doi.org/10.1002/btpr.166

Sin, G., Meyer, A. S., & Gernaey, K. V. (2010). Assessing reliability of cellulose hydrolysis models to support biofuel process design{\textemdash}Identifiability and uncertainty analysis. *Computers*

\& *Chemical Engineering*, *34*(9), 1385–1392. http://doi.org/10.1016/j.compchemeng.2010.02.012

Sin, G., Ödman, P., Petersen, N., Lantz, A. E., & Gernaey, K. V. (2008). Matrix notation for efficient development of first-principles models within PAT applications: Integrated modeling of antibiotic production with Streptomyces coelicolor. *Biotechnology and Bioengineering*, *101*(1), 153–171. http://doi.org/10.1002/bit.21869

Sobol', I. M. (1967). Point distribution in a cube and approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki I Matematicheskoi Fiziki*, *7*(4).

Soini, J., Falschlehner, C., Liedert, C., Bernhardt, J., Vuoristo, J., & Neubauer, P. (2008). Norvaline is accumulated after a down-shift of oxygen in Escherichia coli W3110. *Microbial Cell Factories*, *7*(1), 30. http://doi.org/10.1186/1475-2859-7-30

Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., … Steckenreiter, T. (2017). Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. *Chemie-Ingenieur-Technik*, *89*(5), 542–561. http://doi.org/10.1002/cite.201600175

Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., & Striedner, G. (2017). Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnology Journal*, *12*(7), 1600546. http://doi.org/10.1002/biot.201600546

Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. *Biochemical Engineering Journal*, *135*, 49–60. http://doi.org/10.1016/j.bej.2018.03.016

Stocker, E., Toschkoff, G., Sacher, S., & Khinast, J. G. (2014). Use of mechanistic simulations as a quantitative risk-ranking tool within the quality by design framework. *International Journal of Pharmaceutics*, *475*(1), e245–e255. http://doi.org/10.1016/j.ijpharm.2014.08.055

Stocks, S. M. (2013). *Industrial enzyme production for the food and beverage industries: Process scale up and scale down. Microbial Production of Food Ingredients, Enzymes and Nutraceuticals*. Woodhead Publishing Limited. http://doi.org/10.1533/9780857093547.1.144

Sunya, S., Bideaux, C., Molina-Jouve, C., & Gorret, N. (2013). Short-term dynamic behavior of Escherichia coli in response to successive glucose pulses on glucose-limited chemostat cultures. *Journal of Biotechnology*, *164*(4), 531–542. http://doi.org/10.1016/j.jbiotec.2013.01.014

Tabora, J. E., & Domagalski, N. (2017). Multivariate Analysis and Statistics in Pharmaceutical Process Research and Development. *Annual Review of Chemical and Biomolecular Engineering*, *8*(1), 403–426. http://doi.org/10.1146/annurev-chembioeng-060816-101418

Toumi, A., Jürgens, C., Jungo, C., Maier, B. A., Papavasileiou, V., & Petrides, D. P. (2010). Design and optimization of a large scale biopharmaceutical facility using process simulation and scheduling tools. *Pharmaceutical Engineering*, *30*(2), 1–9. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-77953916038&partnerID=40&md5=0404ad6e20a1bfb7a614b3698e073dba

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction Engineering Principles*. Boston, MA: Springer US. http://doi.org/10.1007/978-1-4419-9688-6

Vlaev, D., Mann, R., Lossev, V., Vlaev, S. D., Zahradnik, J., & Seichter, P. (2000). Macro-mixing and streptomyces fradiae modelling oxygen and nutrient segregation in an industrial bioreactor. *Chemical Engineering Research and Design*, *78*(3), 354–362. http://doi.org/10.1205/026387600527473

von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling

in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, *60*, 86–101. http://doi.org/10.1016/j.compchemeng.2013.08.008

Vrábel, P., Van der Lans, R. G. J. M., Van der Schot, F. N., Luyben, K. C. A. M., Xu, B., & Enfors, S. O. (2001). CMA: Integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chemical Engineering Journal*, *84*(3), 463–474. http://doi.org/10.1016/S1385-8947(00)00271-0

Wang, G., Chu, J., Noorman, H., Xia, J., Tang, W., Zhuang, Y., & Zhang, S. (2014). Prelude to rational scale-up of penicillin production: A scale-down study. *Applied Microbiology and Biotechnology*, *98*(6), 2359–2369. http://doi.org/10.1007/s00253-013-5497-2

Wang, G., Tang, W., Xia, J., Chu, J., Noorman, H., & van Gulik, W. M. (2015). Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Engineering in Life Sciences*, *15*(1), 20–29. http://doi.org/10.1002/elsc.201400172

Wang, H., Khan, F., Ahmed, S., & Imtiaz, S. (2016). Dynamic quantitative operational risk assessment of chemical processes. *Chemical Engineering Science*, *142*, 62–78. http://doi.org/10.1016/j.ces.2015.11.034

Wechselberger, P., Seifert, A., & Herwig, C. (2010). PAT method to gather bioprocess parameters in real-time using simple input variables and first principle relationships. *Chemical Engineering Science*, *65*(21), 5734–5746. http://doi.org/10.1016/j.ces.2010.05.002

Wells, G. J., & Ray, W. H. (2005). Methodology for modeling detailed imperfect mixing effects in complex reactors. *AIChE Journal*, *51*(5), 1508–1520. http://doi.org/10.1002/aic.10407

Whitman, W. G. (1923). A preliminary experimental confirmation of the two-film theory of gas absorption. *Chemical and Metallurgical Engineering*, *29*(4), 146–148. http://doi.org/10.1016/0017-9310(62)90032-7

Xu, B., Jahic, M., Blomsten, G., & Enfors, S. O. (1999). Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli. *Applied Microbiology and Biotechnology*, *51*(5), 564–571. http://doi.org/10.1007/s002530051433

Yang, Y., Xia, J., Li, J., Chu, J., Li, L., Wang, Y., … Zhang, S. (2012). A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *Journal of Biotechnology*, *161*(3), 250–256. http://doi.org/10.1016/j.jbiotec.2012.07.007

Zahradník, J., Mann, R., Fialová, M., Vlaev, D., Vlaev, S. D., Lossev, V., & Seichter, P. (2001). A networks-of-zones analysis of mixing and mass transfer in three industrial bioreactors. *Chemical Engineering Science*, *56*(2), 485–492. http://doi.org/10.1016/S0009-2509(00)00252-9

Zavatti, V., Budman, H., Legge, R., & Tamer, M. (2016). Monitoring of an antigen manufacturing process. *Bioprocess and Biosystems Engineering*, *39*(6), 855–869. http://doi.org/10.1007/s00449-016-1565-1

Zhao, L., Fu, H. Y., Zhou, W., & Hu, W. S. (2015). Advances in process monitoring tools for cell culture bioprocesses. *Engineering in Life Sciences*, *15*(5), 459–468. http://doi.org/10.1002/elsc.201500006

Zhao, W., Buffo, A., Alopaeus, V., Han, B., & Louhi-Kultanen, M. (2017). Application of the compartmental model to the gas-liquid precipitation of CO 2 -Ca(OH) 2 aqueous system in a stirred tank. *AIChE Journal*, *63*(1), 378–386. http://doi.org/10.1002/aic.15567

Zou, X., Xia, J., Chu, J., Zhuang, Y., & Zhang, S. (2012). Real-time fluid dynamics investigation and physiological response for erythromycin fermentation scale-up from 50 L to 132 m3 fermenter.

Paper B

A Probabilistic Model-based Soft Sensor to Monitor Lactic Acid Bacteria Fermentations

Robert Spann[a], Christophe Roca[b], David Kold[b], Anna Eliasson Lantz[c], Krist V. Gernaey[a], Gürkan Sin[a]

[a] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] Chr. Hansen A/S

[c] PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark

Published in the Biochemical Engineering Journal, 135 (2018), 49–60.
https://doi.org/10.1016/j.bej.2018.03.016

Abstract
A probabilistic soft sensor based on a mechanistic model was designed to monitor *S. thermophilus* fermentations, and validated with experimental lab-scale data. It considered uncertainties in the initial conditions, on-line measurements, and model parameters by performing Monte Carlo simulations within the monitoring system. It predicted, therefore, the probability distributions of the unmeasured states, such as biomass, lactose, and lactic acid concentrations. To this end, a mechanistic model was developed first, and a statistical parameter estimation was performed in order to assess parameter sensitivities and uncertainties. The model coupled a biokinetic and a mixed weak acid/base model to predict biological variables and chemical variables like the pH, respectively. In the soft sensor, the limited available on-line measurements, namely the quantity of added ammonia and pH, were used to update the model parameters that were then used as input to the mechanistic model. The soft sensor predicted both the current state variables, as well as the future course of the fermentation, e.g. with a relative mean error of the biomass concentration of 8 %. This successful implementation of a process analytical technology monitoring system opens up further opportunities, including for on-line risk-based monitoring and control applications.

1. Introduction

Lactic acid bacteria (LAB) are used as starter cultures in the dairy industry, to produce probiotics, lactic acid, and exopolysaccharides [1,2]. *Streptococcus thermophilus* strains are aerotolerant, homofermentative LAB and traditionally used as fermentation starter cultures for yogurt and cheese production. The bacteria are produced in batch and fed-batch fermentations, and real-time monitoring of the process is needed in order to understand and optimize the production process. However, robust in-line sensors for key process variables, like biomass, substrate, and lactic acid concentrations, are not available in the required concentration range due to the high complexity of the fermentation system [3]. This makes the real-time quantification of key process variables challenging. The process analytical technology (PAT) guidance from the FDA [4] requested already the development of real-time monitoring and control tools. The tools are meant to enhance the on-line monitoring and control capabilities. Hence, process conditions could be adjusted in real time to assure quality requirements, instead of relying solely on the end product quality control. Although the guidelines were originally intended for the (bio-) pharmaceutical industry, they have also been applied in other life science industries like the food industry.

Soft sensors, which utilize the on-line available measurements, are applied to predict the unknown state variables and monitor the fermentation process in real time [5–7]. There are, generally spoken, data-based and model-based soft sensors, whereas also other approaches exist. Chemometric methods like principal component analysis (PCA) may be applied in data-based soft sensors [8]. Model-based soft sensors can for example be based on mechanistic understanding using first principles models (e.g. the mass balance), or empirical models, when the details of the process are not understood sufficiently. The development of first principles models is based on a fundamental process understanding and mechanistic models may be implemented. Even though the development of mechanistic models might be time consuming, we prefer to use mechanistic models since they have many advantages over black-box models, e.g. that they can be reused and applied to multiple processes by updating the model parameters [7].

Soft sensors rely typically on available on-line and at-line measurements, such as pH, conductivity, dissolved oxygen, heat generation, acid/base addition for pH control, and exhaust gas analysis.

Different spectroscopic measurements, e.g. near-infrared [9], multi-wavelength [10], and Raman [11] spectroscopy have also been used beside other methods in data-based soft sensor applications in fermentations. Mears et al. [12] applied a model-based soft sensor for on-line monitoring to a filamentous fungal fermentation at pilot scale using exhaust gas measurements and ammonia addition, and predicted various state variables (biomass, product, dissolved oxygen, k_La among others). However, especially aerotolerant, homofermentative LAB fermentations lack various on- and at-line measurements, such as exhaust gas measurements, and rely solely on conductivity, pH, and base addition measurements.

When developing and applying mechanistic models for bioprocesses it is good modelling practice for PAT applications to analyze the reliability of the model [13,14]. Unfortunately, models describing LAB fermentations rarely provide reported results of identifiability, sensitivity, or uncertainty analysis, e.g. confidence intervals of the estimated parameters. If a model with unreliable parameters is applied as a soft sensor, predictions will be doubtful and the results could lead to questionable interpretations. Furthermore, a deterministic model implementation may lead to a good fit, but does not take the imperfect knowledge, i.e. uncertainties of model parameters and measurements into account. Several studies implemented soft sensors to monitor LAB [15–18], but they did not consider uncertainties in the model structure.

The aim of this study was therefore to design and evaluate a probabilistic model-based soft sensor in order to monitor *S. thermophilus* fermentations. To this end, a mechanistic model was first developed and validated, and then used as soft sensor for monitoring at lab scale. A statistical parameter estimation was performed to analyze parameter uncertainties. The soft sensor comprised a data reconciliation module, a parameter update module and the dynamic model. The data reconciliation and parameter update module updated model parameters based on the available on-line measurements. One major challenge of this study was that only pH and ammonia addition measurements were available on-line, whereas e.g. exhaust gas measurements were not available. The dynamic model consisted of a biokinetic and a chemical model. The biokinetic model described substrate consumption, biomass growth, and lactic acid secretion while the chemical model comprised a mixed weak acid/base system with the purpose to predict the pH. Monte Carlo simulations of the dynamic model were performed within the monitoring system to account for uncertainties in the lactose (substrate) concentration, ammonia addition rate, and model parameters. The output of the monitoring system was consequently a probability distribution of the state variables.

2. Materials and Methods

2.1 Fermentation conditions

Streptococcus thermophilus batch fermentations were performed in 2 L stirred tank bioreactors (Biostat® B, Sartorius AG, Germany) at 300 rpm, 40 °C, and with nitrogen headspace gassing. The pH was controlled by adding 24 % ammonia solution. 10 batch fermentations were performed under different cultivation conditions (initial lactose concentration 20 or 65 g L^{-1}, pH 5.5 – 7.0) and used for the parameter estimation, model validation, and implementation of the monitoring system (see Table 4 in the Results and Discussion section). The pH (EasyFerm Bio VP 225, Hamilton Robotics, Reno, NV) and ammonia addition (balance value) were measured on-line. The fermentation medium contained 20 or 65 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast

extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$.

2.2 Off-line analysis

Sugars and organic acids were quantified from filtered samples (filter pore size: 0.2 µm) in an HPLC system (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA). It was equipped with an Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA) and a refractive index detector (ERC RefractoMax 520), and run with 5 mM H_2SO_4 at a flow rate of 0.6 mL min^{-1} at 50 °C according to suppliers instructions. Samples were diluted 1:4 with 5 mM H_2SO_4 prior to analysis. Dry cell weight was quantified with replicates of 2 mL cell broth, which were taken in sampling tubes, centrifuged, washed with 0.9 % (w/v) NaCl solution, dried at 70 °C for 24 h, and weighted. Ammonia and phosphate were quantified with the cuvette tests LCK302, LCK303, and LCK350 (Hach®, Manchester, Great Britain).

2.3 Biological model

The dynamic biokinetic model described the evolution of the state variables such as biomass, lactose, and lactic acid of the *S. thermophilus* fermentation. The model was based on the global stoichiometric process equation [19] (Eq. 1 – 2). The biomass growth rate was modelled as a function depending on the lag-time (f_{lag}), lactose inhibition and limitation (f_S) [20], lactate inhibition(f_P) [21], and the pH (f_{pH}) [22] (Eq. 3 – 4). It was assumed that only the dissociated form of lactic acid was growth inhibiting under the investigated pH conditions according to the studies of Schepers et al. [22] and Amrane and Prigent [23]. A biomass composition of $CH_{1.95}O_{0.63}N_{0.22}P_{0.02}$ [24] was assumed in the present study.

$$Lactose + Ammonia + Phosphoric\ acid \rightarrow Biomass + Lactic\ acid + Galactose \quad (1)$$

$$q_S \cdot CH_2O + q_{NH} \cdot NH_3 + q_{Ph} \cdot H_3PO_4 \rightarrow q_X \cdot CH_aO_bN_cP_d + q_P \cdot CH_2O + q_{Gal} \cdot CH_2O \quad (2)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot f_{lag} \cdot f_S \cdot f_P \cdot f_{pH} \cdot C_X \quad (3)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot \left(1 - e^{-\frac{t}{t_{lag}}}\right) \cdot \frac{C_S}{C_S + K_s + \frac{C_S^2}{K_I}} \cdot \frac{1}{1 + e^{K_{P,La}(C_{LA} - K_{La1})}} \cdot e^{-\left(\frac{(pH_{opt} - pH)^2}{\sigma_{pH}^2}\right)} \cdot C_X \quad (4)$$

Where K_{La1} was dependent on the pH:

$$K_{La1} = K_{La} \cdot \frac{1}{1 + e^{K_{P,pH1} * (pH - K_{P,pH2})}} \quad (5)$$

An amended Luedeking and Piret equation [25] that takes only the growth dependent lactic acid synthesis into account was used [26]:

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} \tag{6}$$

The lactose consumption is the sum of the biomass growth and the lactic acid synthesis rate considering the secretion of galactose (Y_{gal}) since the studied strain metabolizes only glucose and secretes galactose under the present cultivation conditions:

$$\frac{dC_S}{dt} = -\left(1 + Y_{gal}\right) \cdot \left(\frac{dC_X}{dt} + \frac{dC_P}{dt}\right) \tag{7}$$

A P-controller with a controller gain (K_P) of 10 mol was applied to maintain the pH at the set point value by adding ammonia solution:

$$NH_4OH_{add} = K_P \cdot (pH_{set} - pH) \tag{8}$$

The model was implemented and solved in MATLAB® (The MathWorks®, Natick, MA) using the solver ode15s. ode15s was chosen because the model contains slow (e.g. biomass growth) and fast time constants (mixed weak acid/base model, see below) resulting in a stiff system of differential equations.

2.4 Mixed weak acid/base model

The purpose of the mixed weak acid/base model was to predict the pH during the fermentation. It comprised the dissociation reactions of the charged compounds in the fermentation broth as described in Musvoto et al. [27] (Table 1). The dissociation reactions of ammonia, phosphoric acid, lactic acid, carbonic acid, water, and an unspecified compound Z were considered. Z accounted for the unknown compounds in the fermentation broth, such as amino acids. The pK_a values were derived from Dawson [28] and Loewenthal et al. [29] (Table 2). The activity coefficients were calculated by a modification of the Debye-Hückel theory from Davies [30]:

$$log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3 \cdot I\right) \tag{9}$$

With the ionic strength (I):

$$I = \frac{1}{2}\sum_i z_i^2 C_i \tag{10}$$

The implemented stoichiometric matrix may be found in the Supplementary Material.

Table 1. Kinetics for the mixed weak acid/base model. f_m and f_d are mono- and divalent activity coefficients, respectively; see Loewenthal et al. (1989) and Musvoto et al. (2000).

Reaction	reaction rate vector	apparent equilibrium constant	
		symbol	value
$NH_4^+ \leftrightarrow NH_3 + H^+$	$K'_{r,NH} \cdot K'_{NH} \cdot [NH_4^+] - K'_{r,NH} \cdot [NH_3] \cdot [H^+]$	K'_{NH}	$10^{-pK_{NH}}$
$H_3PO_4 \leftrightarrow H_2PO_4^- + H^+$	$K'_{r,P1} \cdot K'_{P1} \cdot [H_3PO_4] - K'_{r,P1} \cdot [H_2PO_4^-] \cdot [H^+]$	K'_{P1}	$10^{-pK_{P1}}/f_m^2$
$H_2PO_4^- \leftrightarrow HPO_4^{2-} + H^+$	$K'_{r,P2} \cdot K'_{P2} \cdot [H_2PO_4^-] - K'_{r,P2} \cdot [HPO_4^{2-}] \cdot [H^+]$	K'_{P2}	$10^{-pK_{P2}}/f_d$
$H_2CO_3^* \leftrightarrow HCO_3^- + H^+$	$K'_{r,C1} \cdot K'_{C1} \cdot [H_2CO_3^*] - K'_{r,C1} \cdot [HCO_3^-] \cdot [H^+]$	K'_{C1}	$10^{-pK_{C1}}/f_m^2$
$C_3H_6O_3 \leftrightarrow C_3H_5O_3^- + H^+$	$K'_{r,LA} \cdot K'_{LA} \cdot [C_3H_6O_3] - K'_{r,LA} \cdot [C_3H_5O_3^-] \cdot [H^+]$	K'_{LA}	$10^{-pK_{LA}}/f_m^2$
$H_2O \leftrightarrow OH^- + H^+$	$K'_{r,W} \cdot K'_{W} - K'_{r,W} \cdot [OH^-] \cdot [H^+]$	K'_{W}	$10^{-pK_W}/f_m^2$
$ZH^+ \leftrightarrow Z + H^+$	$K'_{r,Z} \cdot K'_{Z} \cdot [ZH^+] - K'_{r,Z} \cdot [Z] \cdot [H^+]$	K'_{Z}	$10^{-pK_Z}/f_m^2$

Table 2. Parameters of the dynamic model of the *S. thermophilus* fermentations.

Symbol	Value	Reference	Uncertainty classification
Biological model			
K_I	164 g L^{-1}	[20]	
K_{La}	19.80 g L^{-1}	parameter estimation	see Table 3
$K_{P,La}$	0.24 L g^{-1}	parameter estimation	see Table 3
$K_{P,pH1}$	20	expert knowledge	
$K_{P,pH2}$	7	expert knowledge	
K_S	0.79 g L^{-1}	[20]	
pH_{opt}	6.39	parameter estimation	see Table 3
t_{lag}	individual parameter estimation		
Y_{gal}	0.69 g g^{-1}	parameter estimation	see Table 3
α	5.19 g g^{-1}	parameter estimation	see Table 3
μ_{max}	2.06 h^{-1}	parameter estimation	
σ_{pH}	1.42	parameter estimation	see Table 3
Mixed weak acid/base model			
$K'_{r,C1}$	10^7 s^{-1}	[27]	uncertainties in the mixed weak
$K'_{r,LA}$	10^7 s^{-1}	[27]	acid/base model are not considered
$K'_{r,NH}$	10^{12} s^{-1}	[27]	because the pH is measured and
$K'_{r,P1}$	10^8 s^{-1}	[27]	used as input for the data
$K'_{r,P2}$	10^{12} s^{-1}	[27]	reconciliation and parameter update
$K'_{r,W}$	10^{10} s^{-1}	[27]	module
$K'_{r,Z}$	10^7 s^{-1}	[27]	
pK_{C1}	$3404.7/(T - 14.8435 + 0.03279 \cdot T)$	[29]	
pK_{LA}	3.86	[28]	
pK_{NH}	$2835.8/(T - 0.6322 + 0.00123 \cdot T)$	[29]	
pK_{P1}	$799.3/(T - 4.5535 + 0.01349 \cdot T)$	[29]	
pK_{P2}	$1979.5/(T - 5.3541 + 0.01984 \cdot T)$	[29]	
pK_W	14	[29]	
pK_Z	9.4	expert knowledge (amino acid mix)	
T	313.16 K	Measurement	
Initial Conditions			
$C_{Gal,t=0}$	0.0 g L^{-1}		
$C_{Glc,t=0}$	0.0 g L^{-1}		
$C_{H^+,t=0}$	dependent on the pH and ionic strength		
$C_{OH^-,t=0}$	dependent on the pH and ionic strength		
$C_{P,t=0}$	0.0 g L^{-1}		
$C_{S,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{S,t=0}$	65 g L^{-1} for the monitoring system		normal distribution σ = 2.2 g L^{-1}
$C_{tCO,t=0}$	$1.002 \cdot 10^{-5}$ $mol\ L^{-1}$		
$C_{tNH,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{tNH,t=0}$	0.005 g L^{-1} for the monitoring system		
$C_{tPh,t=0}$	off-line measurements for the parameter estimation and validation		
$C_{tPh,t=0}$	2 g L^{-1} for the monitoring system		
$C_{tZ,t=0}$	2 mol L^{-1}		
$C_{X,t=0}$	0.025 g L^{-1} for the parameter estimation and validation		
$C_{X,t=0}$	0.025 g L^{-1} for the monitoring system		normal distribution σ = 0.0008 g L^{-1}

2.5 Parameter estimation

The parameter estimation was performed to fit the experimental lactose, biomass, and lactic acid concentration measurements using the maximum likelihood estimation method from Seber and Wild [31]. The model was fitted to five fermentations, which were controlled at different constant

pH (1x pH 5.5, 2x pH 6.0, 1x pH 6.5, and 1x pH 7.0) and were performed with an initial lactose concentration of 65 g L^{-1}. For the parameter estimation, the pH was held constant at the set point in the simulation, and the mixed weak acid/base model was not considered in order to obtain parameter estimates that were independent of the mixed weak/acid base system. The parameter estimation followed the methodology from Sin and Gernaey [32] as described in Spann et al. [33]. Initial parameter estimates were taken from literature [20–22] (Supplementary Table S1). Sensitivity and identifiability analysis were conducted to find an identifiable parameter subset for regression [32]. Once the regression was completed, the confidence intervals of the estimated parameters were derived from a linear approximation method using the Jacobian matrix of the parameter estimation [34].

The parameter estimation was conducted in MATLAB with the nonlinear least-squares solver lsqnonlin. In the objective function, the weighted error of the model predictions was calculated for the three concentrations lactose, biomass, and lactic acid at all measured time points i (Eq. 11). The residuals vector then contained the weighted error vectors of the three states j.

$$Error_i = \left| \frac{\hat{y}_i - y_{meas,i}}{w_j} \right| \tag{11}$$

where w_j is the maximum value of each specific component, here 65 g L^{-1} for lactose, 30 g L^{-1} for lactic acid, and 6 g L^{-1} for biomass. For model simplification purposes, the lag-time parameter, t_{lag}, was described as a pH dependent distribution, in order to account for the different lag-times observed for fermentations having a different pH set point (Eq. 12). This approach simplifies the model complexity significantly and requires the estimation of only 2 parameters, instead of 5 parameters that would have been needed, if t_{lag} was fitted for each fermentation separately.

$$t_{lag} = 2/e^{\frac{\left(pH_{opt\_lag} - pH\right)^2}{\sigma_{lag}^2}} \tag{12}$$

The uncertainty of the estimated parameters was quantified with the relative error (RE) between the standard deviation of the parameter estimate with respect to the estimated mean value:

$$RE_i = \frac{\sigma_{\hat{\theta}_i}}{\hat{\theta}_i} \tag{13}$$

2.6 Initial conditions

The initial conditions for the dynamic model are given in Table 2.

2.7 Assessment of the model fit

The goodness of fit for the model prediction in the model validation procedure and on-line monitoring application was assessed with the root mean sum of squared errors (RMSSE):

$$RMSSE = \sqrt{\frac{1}{n} \sum_{i}^{n} \left(y_{meas,i} - \hat{y}_i\right)^2} \quad (14)$$

3. Framework for the soft sensor

3.1 Design of the probabilistic model-based soft sensor

The objective of the probabilistic model-based soft sensor is to monitor the *S. thermophilus* fermentation. It predicts the probabilistic distribution of the states, such as biomass, lactose, lactic acid, and pH, in real time based on the on-line available ammonia addition and the pH measurements. The soft sensor consists of a data reconciliation module, a parameter update module, and a dynamic process model (Fig. 1). The model parameters μ_{max} and t_{lag} are updated every 5 minutes based on the latest on-line measurements, and the soft sensor predicts both the current value and the future course of the state variables. The current states are saved as initial conditions for the next interval. Monte Carlo simulations of the dynamic model are performed every interval using samples from the input uncertainty domain. To this end, the Latin hypercube sampling technique was used to generate 100 random samples from the input uncertainty domain in which we included uncertainties in the initial conditions, model parameters, and ammonia addition (Table 2). The outcome from the Monte Carlo simulations was a probability distribution of the state variables.

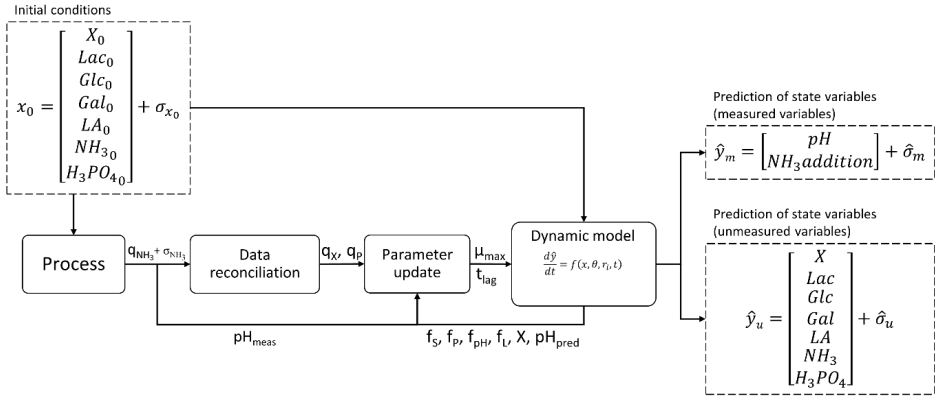


Fig. 1. Block diagram of the probabilistic model-base monitoring system. The initial conditions for the model were defined according to the process specifications including 10 % uncertainties in the lactose and biomass concentration. The on-line measured ammonia addition rate q_{NH3} was used as input for the data reconciliation module to update the biomass growth and lactic acid production rate based on the charge balance and the lactic acid production rate expression. The parameter update module used the updated rates and the pH as input to update the model parameters μ_{max} and t_{lag} for the dynamic model. Monte Carlo simulations of the dynamic model were performed considering uncertainties in the initial lactose concentration, measured ammonia addition, and model parameters. The output of the dynamic model was a probability distribution of the state variables.

3.1.1 Data reconciliation method

The data reconciliation module is based on elemental and bio-energetic conservation principles such as the charge balance (Eq. 15) and the lactic acid production rate expression (Eq. 16). It uses the amount of added ammonia, where one measurement value is available every minute, to update the volumetric biomass growth and lactic acid production rate. The ammonia addition data points of each interval are fitted with a smoothing spline line and the average ammonia addition of the interval, $q_{NH,add}$, is estimated. Missing measurement points can also be handled due to the implementation of the fit. With the data reconciliation the growth rate (q_X) is obtained and used as input for the parameter update module.

$$NH_4^+ + C_3H_5O_3^- = q_{NH,add} + q_P = 0 \tag{15}$$

$$q_P = \alpha \cdot q_X \tag{16}$$

3.1.2 Parameter update

The updated q_X is used to update μ_{max} in every interval. μ_{max} is updated in an iterative procedure until the change of μ_{max} is less than 5 % compared to the previous iteration. In the first iteration (k=1) μ_{max} is calculated based on the updated q_X, the function values, and biomass concentration of the previous interval (Eq. 17). The subsequent iterations use the function values and biomass concentration based on the new μ_{max} value. The function values and biomass concentration derive from an evaluation of the dynamic model.

$$\mu_{max,k} = \frac{q_{X,updated}}{f_{lag,k-1} \cdot f_{S,k-1} \cdot f_{P,k-1} \cdot f_{pH,k-1} \cdot X_{k-1}} \tag{17}$$

It is not possible to use the updated rates (q_P and q_X) directly in the dynamic model, as they resemble only the conditions of the previous 5 minutes. However, inhibition and limitation effects, as well as pH variations, which will occur during a fermentation, influence the rates. It is therefore necessary to calculate the rates within the dynamic model according to Eq. 4 and Eq. 6 in order to predict the future course of the fermentation, as well.

The lag-time parameter t_{lag} is updated based on the measured pH value as soon as the measured pH reaches the controlling value (here pH = 6). The continuous pH measurement is saved every minute. t_{lag} is adjusted so that the modelled and measured pH reach the control value at the same time (Eq. 18). t_{lag} is updated in an iterative procedure until the change is less than 2 % compared to the previous iteration.

$$t_{lag,k} = t_{lag,k-1} + (t_{pH=6,measured} - t_{pH=6,predicted}) \tag{18}$$

Once t_{lag} is updated, the model is run from the beginning, because t_{lag} influences the whole prediction range. The current state is then saved as initial conditions for the next iteration. The updated parameters μ_{max} and t_{lag} are used as input for the dynamic model.

3.1.3 Dynamic mechanistic process model

The dynamic process model comprises the biological model and the mixed weak acid/base model as described in the Materials and Methods section.

3.1.4 Monte Carlo simulations

The Monte Carlo method includes three main steps namely (1) identification and definition of input uncertainties, (2) sampling and (3) Monte Carlo simulation. For step 1, uncertainties in the

biological model parameters, initial lactose concentration, initial biomass concentration, and the ammonia addition are considered (Table 2). The uncertainties of the model parameters are represented by the covariance matrix (includes the standard deviation and correlation matrix), which is derived from the parameter estimation. Uncertainties in the initial lactose and biomass concentration, and the measured ammonia addition are considered to be normally distributed with $3\sigma = 10$ %. The model parameters are assumed to be normally distributed as well because the measurement errors, on which the parameter estimation is founded, are assumed to be normally distributed. In order to account for the ammonia addition uncertainty, samples with a normal distribution, a mean value 1, and $3\sigma = 0.1$ are generated and will be multiplied with the measured ammonia addition rate in the Monte Carlo simulations. Uncertainties in the parameters of the mixed weak acid/base system are not considered since pH predictions were not necessary for the online monitoring application, as pH was directly measured and used as input for the data reconciliation module. The identification of uncertain input sources and the definition of the uncertainty ranges depend on the system studied. In general, this should therefore be systematically evaluated for each studied system separately. In this study, uncertainties of 5 % were expected based on an expert discussion and considering the available data. To be on the safe side, we considered $3\sigma = 10$ % for the uncertainties in the initial conditions and the ammonia measurement.

For step 2, the Latin Hypercube Sampling (LHS) technique is used together with the Iman Conover rank correlation method to induce the correlation matrix in the input domain (see step 1) [35,36]. The sampling procedure features the following generic steps: First, LHS sampling for independent inputs is performed in the unit probability domain [0 1] for N sampling numbers (N =100 used in this study). Then, the correlation matrix is induced via the Iman Conover method [37] for the correlated parameters. Finally, the correlated parameter samples are inverted from the probability domain to real values considering the inverse cumulative distribution function for each input e.g. using the Matlab function icdf. In this study, we assumed both measurement errors as well as parameter estimation errors to be normally distributed hence we set the option “Probability distribution name” to “Normal” in the icdf function. In this step, the user can define any other distribution function deemed appropriate as well (e.g. uniform, gamma etc.).

In step 3, Monte Carlo simulations of the dynamic model are performed for each sample. The output of the Monte Carlo simulations consists of 100 model predictions, representing a probabilistic distribution of the predicted state variables.

4. Results and Discussion

4.1 Parameter estimation

A parameter estimation of the biological model was performed in order to assess the model reliability. Uncertainty and sensitivity analysis were conducted to find an identifiable parameter subset. It must hereby be considered that the estimated parameters depend among others on the nominal parameter values, the cultivation conditions, and the model structure [34]. The first parameter estimation, fitting all biokinetic model parameters, revealed identifiability issues (Supplementary Table S2). $K_{P,pH1}$, $K_{P,pH2}$, K_S, and K_I could not be estimated and were therefore maintained at their initial values for the subsequent steps. The subsequent parameter estimation with the remaining 9 parameters revealed an identifiable parameter subset. The estimated parameter values were in the expected range and in the order of magnitude as known from previous studies (Table 3). It has to be noted that $\mu_{max} = 2.06\ h^{-1}$ was higher than the actual biological value

because it had to compensate for the functions in the growth rate expression. The relative errors of all parameters were lower than 10 %. In addition, all parameters had a significant effect on the model outputs (Supplementary Table S3). Some of the parameters met the criterion of a correlation coefficient smaller than 0.5 to be uniquely identifiable. However, this parameter subset should be considered as a whole due to the linear dependency between most of the parameters. The estimated value of one parameter is conditional on the value of another parameter. Therefore, the covariance matrix of the parameters should be used, e.g. when performing parameter sampling in Monte Carlo simulations, when performing model simulations, as done in this work.

Table 3. Estimated model parameters including the relative error and correlation matrix.

kinetic parameters	estimated parameter value	relative error [%]	correlation matrix								
			μ_{max}	K_P	K_{La}	pH_{opt}	σ_{pH}	α	$pH_{opt-lag}$	σ_{lag}	Y_{gal}
μ_{max}	2.06	1	1	-0.74	-0.83	-0.17	0.5	-0.08	-0.52	0.53	0
$K_{P,La}$	0.24	13		1	0.77	-0.2	-0.58	-0.05	0.39	-0.28	0.04
K_{La}	19.80	0			1	-0.07	-0.54	0.31	0.44	-0.35	-0.28
pH_{opt}	6.39	1				1	-0.52	-0.13	0.76	-0.86	0.02
σ_{pH}	1.42	3					1	0.08	-0.93	0.85	0.06
α	5.19	0						1	-0.1	0.12	-0.4
$pH_{opt-lag}$	5.70	1							1	-0.97	-0.03
σ_{lag}	0.3	9								1	0.03
Y_{gal}	0.69	5									1

The model showed an acceptable fit of the fermentation data (Fig. 2, Supplementary Fig. S1 – 4). To measure process performance, the focus was on the biomass concentration because the cells were the target fermentation product. The RMSSE for biomass was around 0.4 g L^{-1} for many of the fermentations, corresponding to a discrepancy of less than 10 %, giving evidence of a good fit (Table 4). A good model fit was achieved for the fermentations at pH 5.5, 6.0, and 6.5 but not at pH = 7.0, which had an error of 30 %. Furthermore, the secretion of galactose was underestimated in all fermentations. This could be attributed to an inconsistent carbon balance in the experimental fermentation data, where more carbon was produced than lactose consumed. The supplemented yeast extract, which was not taken into account in the model, does contain approximately 6 g L^{-1} carbon [19] when assuming the elemental composition of a *S. cerevisiae* cell for the yeast extract. Hence, amino acids that derived from the yeast extract and were taken up by the cells might have led to the inconsistency in the carbon balance. The parameter estimation aimed therefore not to fit the galactose concentration, but it was anyhow kept in case the model will be extended in future studies.

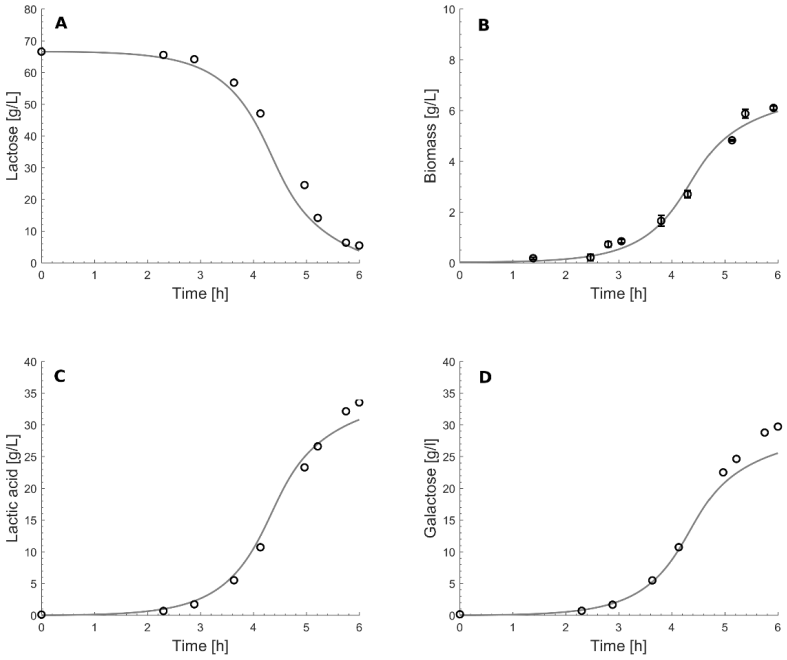


Fig. 2. Model predictions for a *S. thermophilus* lab-scale batch fermentation. . Lactose (A), biomass with standard deviation (B), lactic acid (C), and galactose (D) concentrations. The fermentation was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. The model prediction (solid line) for the measurements (circles) of one of the five lab-scale batches that were used for the parameter estimation is shown. The biomass measurement is shown with the standard deviation.

Table 4. Fermentation conditions and RMSSE of the biomass prediction for all used data sets: parameter estimation, validation, and the monitoring system.

batch data	pH	initial lactose conc. [g L^{-1}]	RMSSE [g L^{-1}]	final biomass [g L^{-1}]*
parameter estimation	5.5	65	0.3	2.45 ± 0.1
	6	65	0.2	6.0 ± 0.2
	6	65	0.6	6.0 ± 0.1
	6.5	65	0.3	5.7 ± 0.1
	7	65	0.6	2.2 ± 0.2
validation	6	20	0.2	2.1 ± 0.2
	6	20	0.2	1.8 ± 0.1
monitoring system	6**	65	0.5 ± 0.1	5.9 ± 0.1
	6**	65	0.5 ± 0.1	5.8 ± 0.2
	6**	65	0.5 ± 0.1	5.9 ± 0.2

* with standard deviation of the last measurement at ca. 6 h fermentation time
** the initial pH was the pH of the medium (around pH = 7). The fermentation was controlled at pH = 6.

The evaluation of the pH function f_{pH} showed a clear maximum at pH = 6.4 (Fig. 3A). Furthermore, growth was already reduced by 25 % at pH = 5.5 and 7. Similar trends of the influence of the pH on the growth of LAB were observed in other studies [21,38,39]. These studies found slightly different pH optimums in the range between 6 and 7 since different strains were studied. The growth inhibition by lactate was pH dependent, as well (Fig. 3B). 20 g L^{-1} lactate inhibited growth by 50 % in the pH range from 5.5 to 6.5, whereas at pH = 7 already 10 g L^{-1} lactate inhibited growth by 50 %. pH dependent inhibition of growth caused by lactate was also already observed for the lactic acid producing bacterium *Enterococcus faecalis* [40]. This underlines the necessity of the pH dependent lactate inhibition parameter K_{La1} (Eq. 5).

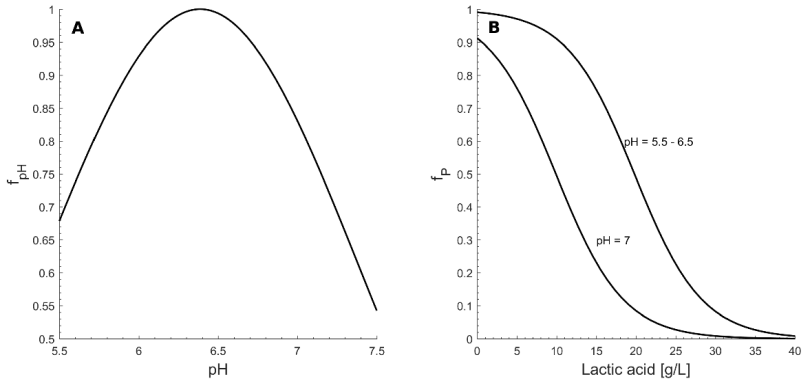


Fig. 3. Growth affecting functions of pH and lactate inhibition. pH function vs. pH (A) and lactate inhibition function vs. lactic acid concentration (B).

4.2 Model validation

Following the statistical assessment of the quality of the parameter estimates above, the model was validated against two independent fermentation data sets, which were performed at pH = 6.0 and an initial lactose concentration of 20 g L^{-1} (Fig. 4, Supplementary Fig. S5). The model predicted the measured lactose, biomass, lactic acid, and galactose concentrations (Fig. 4 A-D). The lag-time parameter t_{lag} was fitted for both fermentations because the lag time differs from batch to batch. The assessment of the validation model fit showed that the model gave an acceptable prediction accuracy with an RMSSE for biomass of 0.2 g L^{-1} (Table 4). The pH prediction was also very accurate with a discrepancy of less than 0.1 pH units (Fig. 4 E). In the beginning of the fermentation, the pH dropped from 6.1 to 6.0 before the controller started to add ammonia solution. This drop was predicted to be faster than actually measured, which could be attributed to a slightly different buffer capacity of the medium in reality compared to the mixed weak acid/base model. However, a prediction accuracy within ± 0.1 pH units was deemed sufficiently accurate for monitoring purposes, as pH measurement errors were expected to be in the same range. The only exception of an accurate pH fit was at the moment when the pH controller started: too much ammonia was added in the experiment so that the pH showed an overshoot. The pH prediction is closely dependent on the predictions of the ammonia addition and lactic acid concentration. The validity of the mixed weak acid/base model was therefore demonstrated by a correct prediction of the added ammonia solution (Fig. 4 F), as the pH is held constant by adding ammonia solution. Nevertheless, the validity of the applied Davis equation to calculate the activity coefficients (Eq. 9) for $I \leq 0.5\ mol\ L^{-1}$ has to be noted, and could be improved in future studies in particular for fermentations with an ionic strength higher than 1 mol L^{-1}. Overall, these results indicate the validity of the model, which encourages its further application for monitoring of a fermentation process as presented below.

4.3 Application of the probabilistic model-based soft sensor

The probabilistic model-based soft sensor was applied to the data sets of three historical fermentations, where the historical on-line data was used as it would be available on-line. Here, the initial pH was around 7 and the pH dropped to the control value 6 due to acid secretion as by-product during the LAB fermentation (Fig. 5, Supplementary Fig. S6 and S7, while the Supplementary Movies show the virtual on-line implementation of the soft sensor). The on-line data, namely pH and quantity of added ammonia were used as an input to update the monitoring system (Fig. 5 left column). The Monte Carlo propagation of the error for the biomass, lactose, and lactic acid concentration is then predicted by the monitoring system (Fig. 5 middle and right column). The predictions of the future states of the system are shown at different times, 2 h, 2 h 40 min, 3 h, 4 h, and 6 h (Fig. 5 rows). Since the pH was higher than the control value 6, no ammonia solution was added within the first 2 h and 35 min. Therefore, no data reconciliation and parameter update were conducted (Fig. 5, time =2 h), and Monte Carlo simulations were performed in the defined input uncertainty space (Table 2 and Supplementary Fig. S8) considering uncertainties in the biological model parameters and initial conditions. The monitoring system was running without updating the parameters until the ammonia addition started to control the pH (Fig. 5, at time = 2 h 40 min). At this point, t_{lag} was updated ensuring that the pH controller in the experiment and simulation started at the same time. It is clear that there is a lag-time variation from batch to batch, which has to be taken into account. On the one hand, a dependency on the pH measurement is introduced by this procedure. On the other hand, it is the only possibility – given the limited amount of on-line measurements – to align the modelled and measured ammonia addition, which is crucial for the monitoring system. Once t_{lag} was updated, and the ammonia

addition started, the data reconciliation and parameter update modules updated μ_{max} every 5 minutes, as described in the Framework description. With time, more measurement information was available and the prediction accuracy of the state variables increased (Fig.5, time =3 h – 6 h).

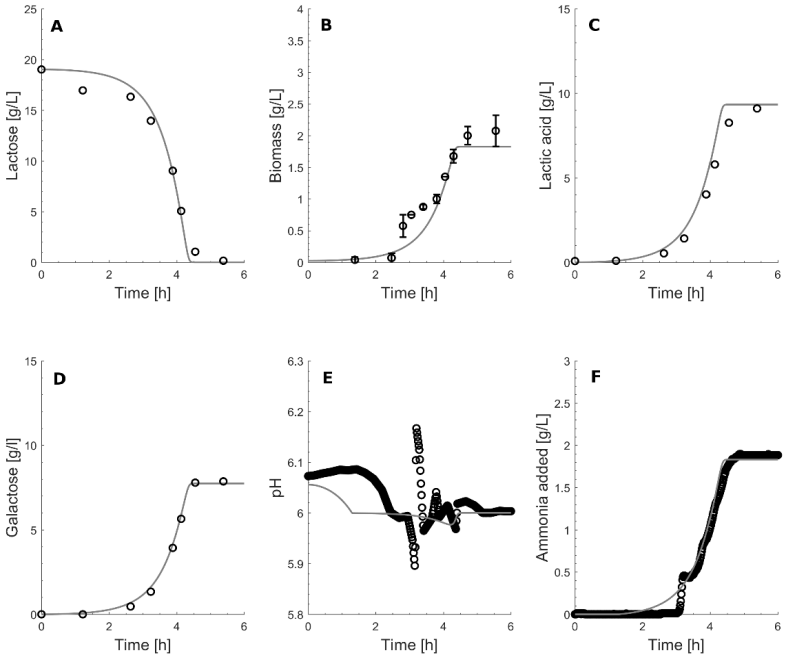


Fig. 4. Model prediction for a validation lab-scale batch fermentation. Lactose (A), biomass with standard deviation (B), lactic acid (C), galactose (D) concentrations, pH (E), and the added ammonia amount (F). The *S. thermophilus* fermentation was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. Model prediction (solid line) for the measurements (circles) of one of the two validation lab-scale batches.

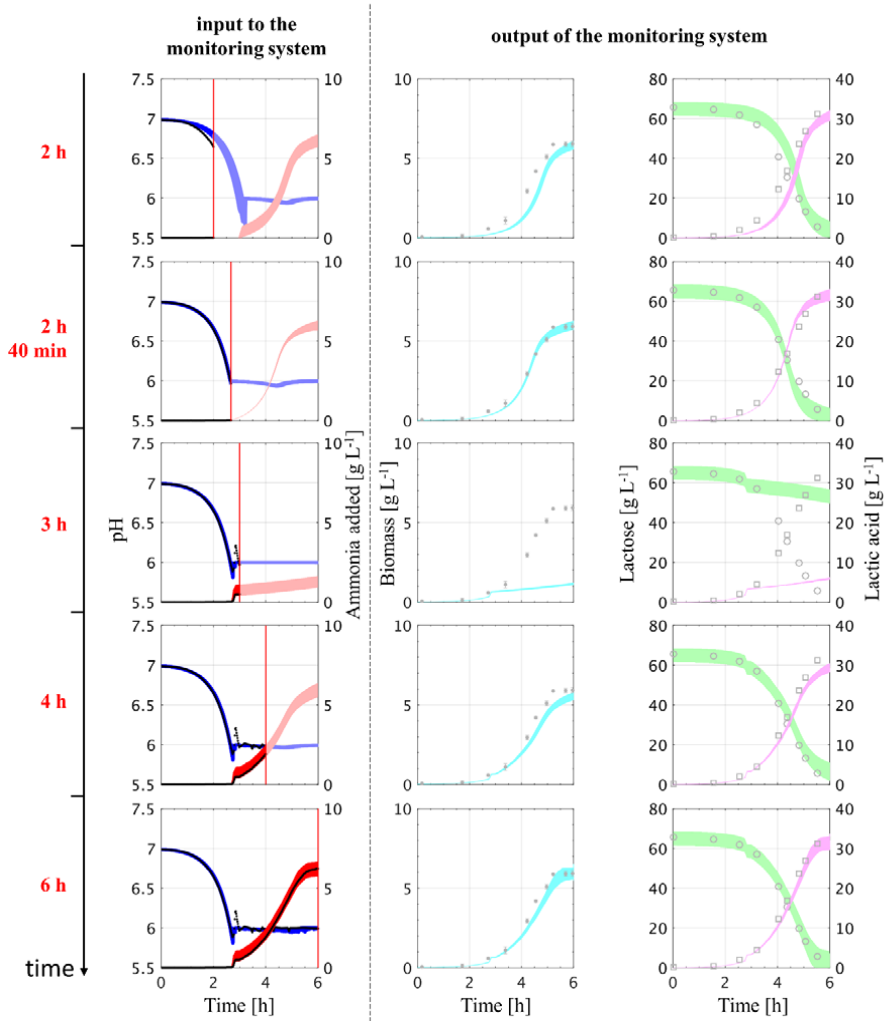


Fig. 5. Probabilistic monitoring system applied to lab-scale batch data of a *S. thermophilus* fermentation. The monitoring system reads in the on-line available data (left column, black dots), ammonia addition and pH, and predicts the state variables every 5 minutes (middle and right column). 100 Monte Carlo simulations of the dynamic model were performed within the monitoring system considering uncertainties in the initial conditions, ammonia addition, and model parameters. The 95 % confidence intervals of the predictions are shown at five time points during the fermentation (2 h, 2 h 40 min, 3 h, 4 h, 6 h). Predictions of the pH (blue), ammonia addition (red), biomass (cyan), lactose (green), and lactic acid (magenta) concentrations are shown. The off-line measurements for biomass (gray dot with standard deviation), lactose (gray circle), and lactic acid (gray square) are shown for comparison only, but were not used for the data reconciliation and parameter update (see Fig. 1).

The RMSSE for biomass was 0.8 g L^{-1} when the fermentation started, and improved to 0.5 g L^{-1} at the end of the fermentation (both with a standard deviation of 0.1 g L^{-1}) (Fig. 6). Mainly, the update of t_{lag} after 2 h and 40 minutes of the fermentation improved the prediction accuracy. The reproducibility of the Monte Carlo simulations was validated as the RMSSE for biomass varied less than 0.5 % in 10 repeated Monte Carlo simulations with 100 input samples in each simulation. Changing the tolerance limit to estimate μ_{max} in the iterative update procedure (Eq. 17) to 1% and 0.1% did not improve the prediction accuracy for the presented fermentations (data not shown). However, this might be necessary for other applications. In summary, an accurate prediction of the state variables was achieved.

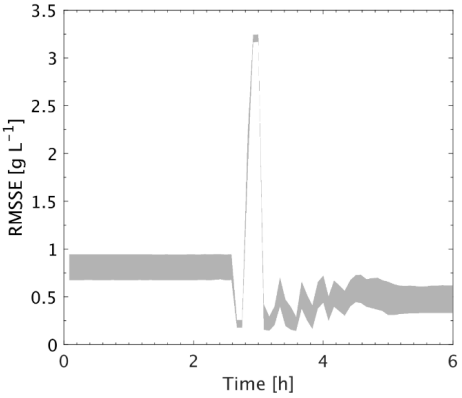


Fig. 6. 95% confidence interval of the RMSSE for the biomass prediction during the probabilistic monitoring of a *S. thermophilus* fermentation.

Several reports have implemented soft sensors to monitor LAB fermentations. Acuña et al. [16] and Peter and Röck [15] implemented a model-based monitoring system for LAB fermentations using the base addition and pH measurements, whereas the second implementation is limited to monitoring the lactic acid concentration. Fayolle et al. [17] and Payot et al. [18] designed a data-based soft sensor using mid-infrared spectroscopy and conductivity, respectively. However, all studies presented deterministic predictions and did not consider the imperfect knowledge of the process by taking uncertainties into account.

Contrary to the earlier published investigations, this study accounted for several sources of uncertainties in the probabilistic monitoring system and assessed the combined effect of system uncertainties on the predictions. The initial conditions, on-line measurements, and biological model parameters were considered as uncertainty sources. The concentrations of the medium components (initial conditions) vary from batch to batch as the medium preparation procedure underlies uncertainties. The biomass concentration depends on the cryo-stock and pre-culture quality. Since the monitoring system relies on the ammonia addition measurement, it is important to incorporate measurement uncertainties, as well. Mears et al. [12] pointed out that an error of 5 % of the carbon evolution rate or oxygen uptake rate, caused by measurement errors in the exhaust gas, led to errors of more than 50 % in the model prediction of the final biomass concentration in a filamentous fungi process at pilot scale. The exact extent of uncertainties of the

initial conditions and measurements could not be determined in the present study because statistically relevant data was not available. The implemented uncertainties were instead based on expert knowledge. However, the model parameter uncertainties were obtained in the parameter regression step that has been presented above.

The monitoring system predicted the current state variables and forecasted the future course of the fermentation and could therefore support a lean production. If this monitoring system is applied at production sites, it will provide plant operators with a PAT tool to monitor the biological variables in the fermentation process , such as biomass concentration, instead of on-line ammonia addition measurements, where the latter are difficult to comprehend (as shown in the virtual implementation in the Supplementary Movies). In addition, the tool could predict whether and when the target cell yield will be reached. This helps run the batch period optimally and schedule other unit operations: All downstream processing steps could be coordinated with the upstream fermentation batch time and therefore be prepared in time. Moreover, cleaning, sterilization, media preparation, and pre-culture steps affiliated with the start up of the batch process could be optimized to reduce the overall downtime of the fermentation unit. An optimized schedule with efficient utilization of the different operation units can contribute to a more economical operation of the production plant. The monitoring system could also cover the early diagnosis of process failures and warn the operators if biomass growth had stopped unexpectedly. The standstill of ammonia addition is a sign of interrupted biomass growth, as it happened in the shown fermentation after 3 h (Fig. 5, time = 3 h). However, a warning should only be given in case the ammonia addition stopped for a longer period of time – in contrast to the present fermentation, where the ammonia addition stopped only for a short while because the pH controller overshot. Furthermore, the system could be extended to calculate the risk of not achieving the target biomass yield as a result of the outcome of the soft sensor, which provides the probability distribution of biomass concentration at the end of a given batch. It could then be implemented for risk-based monitoring and be further developed for control purposes.

5. Conclusion

A probabilistic model-based soft sensor was proposed for the monitoring of *S. thermophilus* fermentations. State variables, such as biomass, substrate, and lactic acid, which were not possible to be measured on-line, could be successfully predicted. The predictions were based on very limited available on-line measurements, namely base addition and pH measurements since exhaust gas measurements were not available. The aim was achieved by coupling a biokinetic model and a mixed weak acid/base model (for the pH calculation), which were validated comprehensively. Uncertainties in the initial substrate concentration, base addition, and biological model parameters were quantified and accounted for using Monte Carlo simulations in the probabilistic monitoring system. The future objective of this study will be to implement the monitoring system for on-line risk-based monitoring and control in pilot- and large-scale LAB studies.

Competing interest

The authors declare that they have no competing interests.

Acknowledgement
This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project). We are thankful for the cooperation with Chr. Hansen A/S, and the support from Klaus Pellicer Alborch (Technische Universität Berlin) during the experiments.

Nomenclature

C_{Gal}	galactose concentration (g L^{-1})
C_{Glc}	glucose concentration (g L^{-1})
C_{H^+}	H^+ concentration (mol L^{-1})
C_{LA}	lactate concentration (g L^{-1})
C_{OH^-}	OH^- concentration (mol L^{-1})
C_P	total lactic acid (lactate and lactic acid) concentration (g L^{-1})
C_S	lactose (substrate) concentration (g L^{-1})
C_{tCO}	total carbonic acid ($H_2CO_3^*$ and HCO_3^-) concentration (mol L^{-1})
C_{tNH}	total concentration of NH_4^+ and NH_3 (g L^{-1})
C_{tPh}	total concentration of H_3PO_4, $H_2PO_4^-$, and HPO_4^{2-}(g L^{-1})
C_{tZ}	total concentration of the unknown compound (dissociated and undissociated form) (mol L^{-1})
C_X	biomass concentration (g L^{-1})
$Error_i$	Weighted model prediction error of a component at time point i
f_d	divalent activity coefficients (-)
f_{lag}	lag-time function (-)
f_m	monovalent activity coefficients (-)
f_P	lactic acid inhibition function (-)
f_{pH}	pH dependency function (-)
f_S	substrate limitation and inhibition function (-)
$H_2CO_3^*$	dissolved CO_2 and H_2CO_3
I	ionic strength (g L^{-1})
K'_{C1}	apparent equilibrium constant for the carbonic acid system (-)
K_I	substrate inhibition parameter (g L^{-1})
K_{La}	lactate inhibition parameter (g L^{-1})
K_{La1}	pH dependent lactate inhibition parameter (g L^{-1})
K'_{LA}	apparent equilibrium constant for the lactic acid system (-)
K'_{NH}	apparent equilibrium constant for the ammonia system (-)
K_P	P-controller controller gain
$K_{P,La}$	2. lactate inhibition parameter (L g^{-1})
$K_{P,pH1}$	lactate inhibition pH parameter (-)
$K_{P,pH2}$	2. lactate inhibition pH parameter (-)
K'_{P1}	apparent equilibrium constant for the phosphoric acid system (-)
K'_{P2}	apparent equilibrium constant for the dihydrogen phosphate system (-)
$K'_{r,C1}$	apparent reverse rate constant for carbonic acid dissociation (s^{-1})
$K'_{r,LA}$	apparent reverse rate constant for lactic acid dissociation (s^{-1})
$K'_{r,NH}$	apparent reverse rate constant for NH_4 dissociation (s^{-1})
$K'_{r,P1}$	apparent reverse rate constant for H_3PO_4 dissociation (s^{-1})
$K'_{r,P2}$	apparent reverse rate constant for $H_2PO_4^-$ dissociation (s^{-1})
$K'_{r,W}$	apparent reverse rate constant for water dissociation (s^{-1})
K_S	substrate limitation parameter (g L^{-1})
K'_W	apparent equilibrium constant for the water system (-)
K'_Z	apparent equilibrium constant for the unspecified compound system (-)
n	number of measurement points
pH_{opt}	optimal pH parameter in the pH function (-)
pH_{opt_lag}	optimal pH for the lag-time fit (-)
pH_{set}	pH control set point (-)

pK_{C1}	pK_a constant for carbonic acid dissociation
pK_{LA}	pK_a constant for lactic acid dissociation
pK_{NH}	pK_a constant for NH_4 dissociation
pK_{P1}	pK_a constant for H_3PO_4 dissociation
pK_{P2}	pK_a constant for $H_2PO_4^-$ dissociation
pK_W	pK_a constant for water dissociation
pK_Z	pK_a constant for the unspecified compound dissociation
q_{Gal}	volumetric galactose secretion rate (C-mol L^{-1} h^{-1})
q_{NH}	volumetric ammonia consumption rate (mol L^{-1} h^{-1})
$q_{NH,add}$	volumetric ammonia addition rate (mol L^{-1} h^{-1})
q_P	volumetric lactic acid secretion rate (C-mol L^{-1} h^{-1})
q_{Ph}	volumetric phosphoric acid consumption rate (mol L^{-1} h^{-1})
q_S	volumetric substrate consumption rate (C-mol L^{-1} h^{-1})
q_X	volumetric biomass growth rate (C-mol L^{-1} h^{-1})
RE	relative error (-)
RMSSE	root mean sum of squared errors (g L^{-1})
T	temperature in the bioreactor (K)
t	time variable (h)
t_{lag}	lag-time coefficient (h)
w_j	maximum value of the state j for the weighted error in the objective function
Y_{gal}	galactose yield (g g^{-1})
z_i	charge number of the i-th ion
$\hat{y}_i$	i-th model value of one output (g L^{-1})
$y_{meas,i}$	i-th measurement value of one output (g L^{-1})

Greek Letters

α	growth related production coefficient of lactic acid (g g^{-1})
$\hat{\theta}_i$	estimated parameter value
μ_{max}	maximum specific growth rate (h^{-1})
σ	standard deviation
σ_{pH}	spread parameter is the gaussian pH function
σ_{lag}	standard deviation of the lag-time fit
$\sigma_{\hat{\theta}_i}$	standard deviation of the estimated parameter

References

[1] A. Von Wright, L. Axelsson, Lactic Acid Bacteria: An Introduction, in: S. Lahtinen, A.C. Ouwehand, S. Salminen, A. VonWright (Eds.), Lact. Acid Bact. Microbiol. Funct. Asp., 4th ed., CRC Press, Taylor & Francis, Boca Raton, 2012: pp. 1–16.

[2] S. Patel, A. Majumder, A. Goyal, Potentials of Exopolysaccharides from Lactic Acid Bacteria, Indian J. Microbiol. 52 (2012) 3–12.

[3] B. Sonnleitner, Automated Measurement and Monitoring of Bioprocesses: Key Elements of the M3C Strategy, in: Adv. Biochem. Eng. Biotechnol., 2012: pp. 1–33.

[4] FDA, Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, U.S. Food and Drug Administration, U.S. Department of Health and Human Services, 2004.

[5] R. Luttmann, D.G. Bracewell, G. Cornelissen, K. V. Gernaey, J. Glassey, V.C. Hass, C. Kaiser, C. Preusse, G. Striedner, C.F. Mandenius, Soft sensors in bioprocessing: A status report and recommendations, Biotechnol. J. 7 (2012) 1040–1048.

[6] K. V Gernaey, A.E. Lantz, P. Tufvesson, J.M. Woodley, G. Sin, Application of mechanistic models to fermentation and biocatalysis for next-generation processes., Trends Biotechnol. 28 (2010) 346–54.

[7] L. Mears, S.M. Stocks, M.O. Albaek, G. Sin, K. V. Gernaey, Mechanistic Fermentation Models for Process Design, Monitoring, and Control, Trends Biotechnol. 35 (2017) 914–924.
[8] P. Kadlec, B. Gabrys, S. Strandt, Data-driven Soft Sensors in the process industry, Comput. Chem. Eng. 33 (2009) 795–814.
[9] J. Alves-Rausch, R. Bienert, C. Grimm, D. Bergmaier, Real time in-line monitoring of large scale Bacillus fermentations with near-infrared spectroscopy, J. Biotechnol. 189 (2014) 120–128.
[10] P. Ödman, C.L. Johansen, L. Olsson, K. V. Gernaey, A.E. Lantz, On-line estimation of biomass, glucose and ethanol in Saccharomyces cerevisiae cultivations using in-situ multi-wavelength fluorescence and software sensors, J. Biotechnol. 144 (2009) 102–112.
[11] A. Golabgir, C. Herwig, Combining Mechanistic Modeling and Raman Spectroscopy for Real-Time Monitoring of Fed-Batch Penicillin Production, Chemie-Ingenieur-Technik. 88 (2016) 764–776.
[12] L. Mears, S.M. Stocks, M.O. Albaek, G. Sin, K. V. Gernaey, Application of a mechanistic model as a tool for on-line monitoring of pilot scale filamentous fungal fermentation processes—The importance of evaporation effects, Biotechnol. Bioeng. 114 (2017) 589–599.
[13] R. Lencastre Fernandes, V.K. Bodla, M. Carlquist, A.-L. Heins, A. Eliasson Lantz, G. Sin, K. V. Gernaey, Applying Mechanistic Models in Bioprocess Development, in: Adv. Biochem. Eng. Biotechnol., 2013: pp. 137–166.
[14] G. Sin, K. V Gernaey, A.E. Lantz, Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis., Biotechnol. Prog. 25 (2009) 1043–1053.
[15] E. Peter, H. Röck, On-line estimation of lactic acid concentration during batch fermentations of Streptococcus thermophilus based on pH measurement, 12th Int. Conf. Control. Autom. Syst. (2012) 851–855.
[16] G. Acuña, E. Latrille, C. Béal, G. Corrieu, A. Chéruy, On-line estimation of biological variables during pH controlled lactic acid fermentations, Biotechnol. Bioeng. 44 (1994) 1168–1176.
[17] P. Fayolle, D. Picque, G. Corrieu, Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy, Vib. Spectrosc. 14 (1997) 247–252.
[18] T. Payot, M. Fick, G. Chimique, On-line estimation of lactic acid concentration by conductivity measurement in fermentation broth, Biotechnol. Tech. 11 (1997) 17–20.
[19] J. Villadsen, J. Nielsen, G. Lidén, Bioreaction Engineering Principles, Springer US, Boston, MA, 2011.
[20] C. Åkerberg, K. Hofvendahl, B. Hahn-Hägerdal, G. Zacchi, Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour, Appl. Microbiol. Biotechnol. 49 (1998) 682–690.
[21] M. Aghababaie, M. Khanahmadi, M. Beheshti, Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures, Food Bioprod. Process. 94 (2015) 657–667.
[22] A.W. Schepers, J. Thibault, C. Lacroix, Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation, Enzyme Microb. Technol. 30 (2002) 187–

194.
[23] A. Amrane, Y. Prigent, Influence of an initial addition of lactic acid on growth, acid production and their coupling for batch cultures of Lactobacillus helveticus., Bioprocess Eng. 19 (1998) 307–312.
[24] A.P. Oliveira, J. Nielsen, J. Förster, Modeling Lactococcus lactis using a genome-scale flux model., BMC Microbiol. 5 (2005) 39.
[25] R. Luedeking, E.L. Piret, A Kinetic Study of the Lactic Acid Fermentation. Batch Process at Controlled pH, J. Biochem. Microbiol. Technol. Eng. 1 (1959) 393–412.
[26] R.Y. Peng, T.C.K. Yang, H. Wang, Y. Lin, C. Cheng, Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model, J. Chinese Agric. Chem. Soc. 35 (1997) 485–494.
[27] E.V. Musvoto, M.C. Wentzel, R.E. Loewenthal, G.A. Ekama, Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems, Water Res. 34 (2000) 1857–1867.
[28] R.M.C. Dawson, Data for Biochemical Research, Clarendon Press, Oxford, 1969.
[29] R.E. Loewenthal, G.A. Ekama, G.R. Marais, Mixed weak acid/base systems Part I - Mixture characterisation., Water SA. 15 (1989) 3–24.
[30] C.W. Davies, Ion Association, Butterworth, Londond, 1962.
[31] G.A.F. Seber, C.J. Wild, Nonlinear Regression, (1989) 792.
[32] G. Sin, K. V. Gernaey, Data Handling and Parameter Estimation, in: M.C.M. van Lossdrecht, P.H. Nielsen, C.M. Lopez-Vazquez, D. Brdjanovic (Eds.), Exp. Methods Wastewater Treat., IWA Publishing Company, 2016: pp. 201–234.
[33] R. Spann, C. Roca, D. Kold, A. Eliasson Lantz, K. V. Gernaey, G. Sin, A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model, in: A. Espuña, M. Graelss, L. Puigjaner (Eds.), Proc. 27th Eur. Symp. Comput. Aided Process Eng. – ESCAPE 27, Elsevier, 2017: p. 3042.
[34] G. Sin, A.S. Meyer, K. V. Gernaey, Assessing reliability of cellulose hydrolysis models to support biofuel process design—Identifiability and uncertainty analysis, Comput. Chem. Eng. 34 (2010) 1385–1392.
[35] M.D. McKay, R.J. Beckman, W.J. Conover, Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics. 21 (1979) 239–245.
[36] G. Sin, K. V. Gernaey, M.B. Neumann, M.C.M. van Loosdrecht, W. Gujer, Uncertainty analysis in WWTP model applications: A critical discussion using an example from design, Water Res. 43 (2009) 2894–2906.
[37] R.L. Iman, W.J. Conover, A distribution-free approach to inducing rank correlation among input variables, Commun. Stat. - Simul. Comput. 11 (1982) 311–334.
[38] M.-P.P. Zacharof, R.W. Lovitt, Modelling and simulation of cell growth dynamics, substrate consumption, and lactic acid production kinetics of Lactococcus lactis, Biotechnol. Bioprocess Eng. 18 (2013) 52–64.
[39] H. Ohara, K. Hiyama, T. Yoshida, Kinetic study on pH dependence of growth and death of Streptococcus faecalis, Appl. Microbiol. Biotechnol. 38 (1992) 403–407.
[40] A.D. Nandasana, S. Kumar, Kinetic modeling of lactic acid production from molasses using Enterococcus faecalis RKY1, Biochem. Eng. J. 38 (2008) 277–284.

Supplementary material

Table S1. Initial parameter values.

kinetic parameters	initial value	source of initial value	lower bound	upper bound
μ_{max}	1.18 h^{-1}	[21]	1	5
K_S	0.79 g L^{-1}	[20]	0.5	5
K_I	164 g L^{-1}	[20]	50	200
$K_{P,La}$	0.2 L g^{-1}	[22]	0.1	1
K_{La}	25 g L^{-1}	[22]	10	30
pH_{opt}	6	expert knowledge	5.7	6.5
σ_{pH}	1.54	[22]	0.1	3
α	5 g g^{-1}	expert knowledge	0.1	30
K_{P_pH1}	20	expert knowledge	1	50
K_{P_pH2}	7	expert knowledge	6.8	7.2
$pH_{opt\text{-}lag}$	5.7	expert knowledge	5.5	6
σ_{lag}	0.4	expert knowledge	0.3	0.5
Y_{gal}	0.6 g g^{-1}	expert knowledge	0	1

Table S2. First parameter estimation: Estimated parameter values, relative errors, and local sensitivity analysis

kinetic parameters	estimated parameter value	relative error [%]	δ^{msqr} lactose	δ^{msqr} biomass	δ^{msqr} lactate
μ_{max}	2.23	19	57	5	28
K_S	2.88	73	2	0	1
K_I	139.10	1	17	2	8
$K_{P,La}$	0.21	33	6	1	3
K_{La}	20.19	0	23	2	11
pH_{opt}	6.49	6	142	13	70
σ_{pH}	1.60	9	11	1	5
α	5.20	0	8	2	6
K_{P_pH1}	6.96	0	0	0	0
K_{P_pH2}	6.95	3	1	0	0
$pH_{opt\text{-}lag}$	5.69	1	403	38	198
σ_{lag}	0.30	118	22	2	11
Y_{gal}	0.69	5	12	0	0

Table S3. Final parameter estimation: Estimated parameter values, relative errors, and local sensitivity analysis

kinetic parameters	estimated parameter value	relative error [%]	δ^{msqr} lactose	δ^{msqr} biomass	δ^{msqr} lactate
μ_{max}	2.06	1	56	5	27
$K_{P,La}$	0.24	13	5	0	3
K_{La}	19.80	0	24	2	12
pH_{opt}	6.39	1	135	13	66
σ_{pH}	1.42	3	8	1	4
α	5.19	0	8	2	5
$pH_{opt-lag}$	5.70	1	409	39	201
σ_{lag}	0.3	9	22	2	11
Y_{gal}	0.69	5	13	1	0

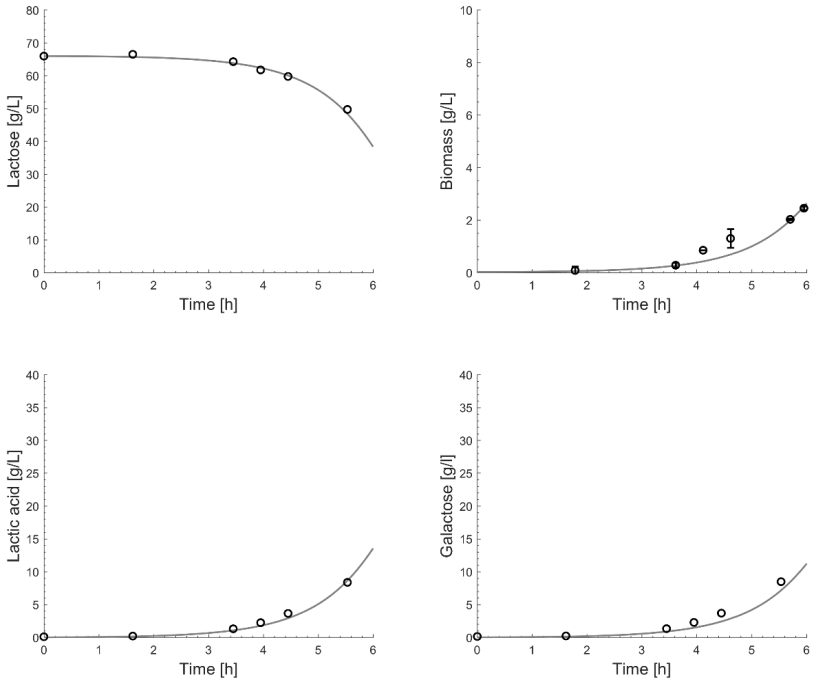


Fig. S1. Model predictions for a *S. thermophilus* lab-scale batch fermentation at pH = 5.5. It was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 5.5. Model prediction (solid line) for the measurements (circles). The biomass measurement is shown with the standard deviation. This fermentation was used for the parameter estimation.

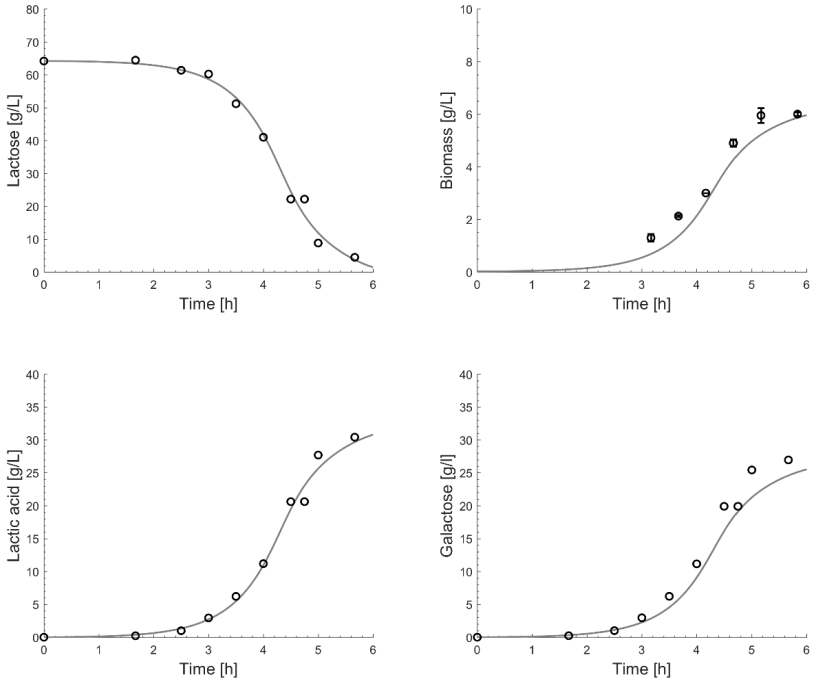


Fig. S2. Model predictions for a *S. thermophilus* lab-scale batch fermentation at pH = 6.0. It was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6.0. Model prediction (solid line) for the measurements (circles). The biomass measurement is shown with the standard deviation. This fermentation was used for the parameter estimation.

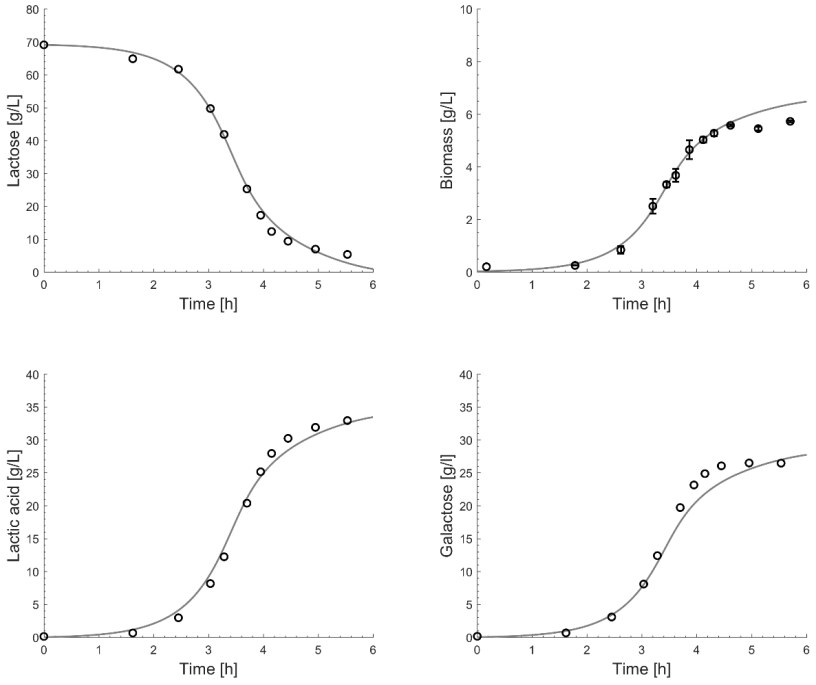


Fig. S3. Model predictions for a *S. thermophilus* lab-scale batch fermentation at pH = 6.5. It was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6.5. Model prediction (solid line) for the measurements (circles). The biomass measurement is shown with the standard deviation. This fermentation was used for the parameter estimation.

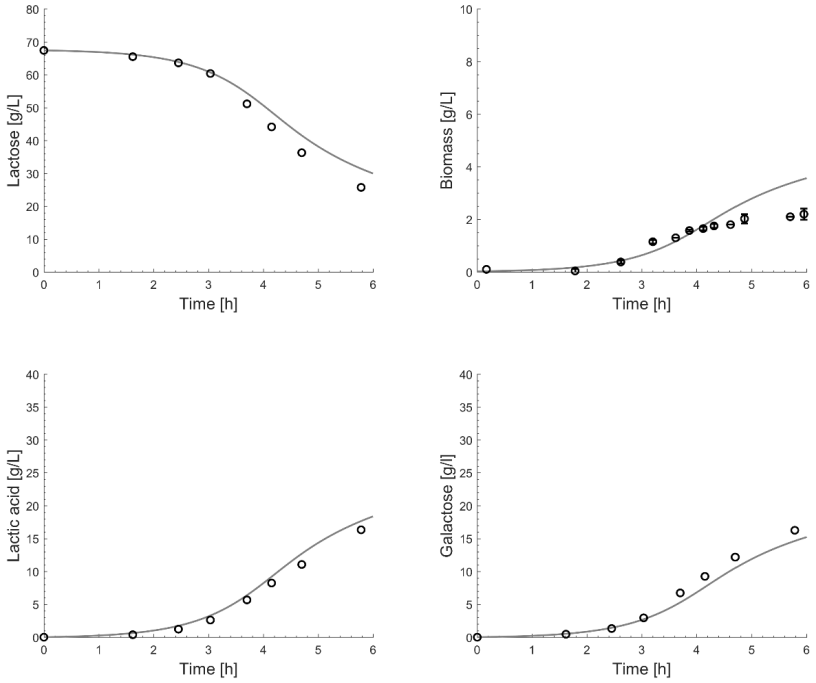


Fig. S4. Model predictions for a *S. thermophilus* lab-scale batch fermentation at pH = 7.0. It was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 7.0. Model prediction (solid line) for the measurements (circles). The biomass measurement is shown with the standard deviation. This fermentation was used for the parameter estimation.

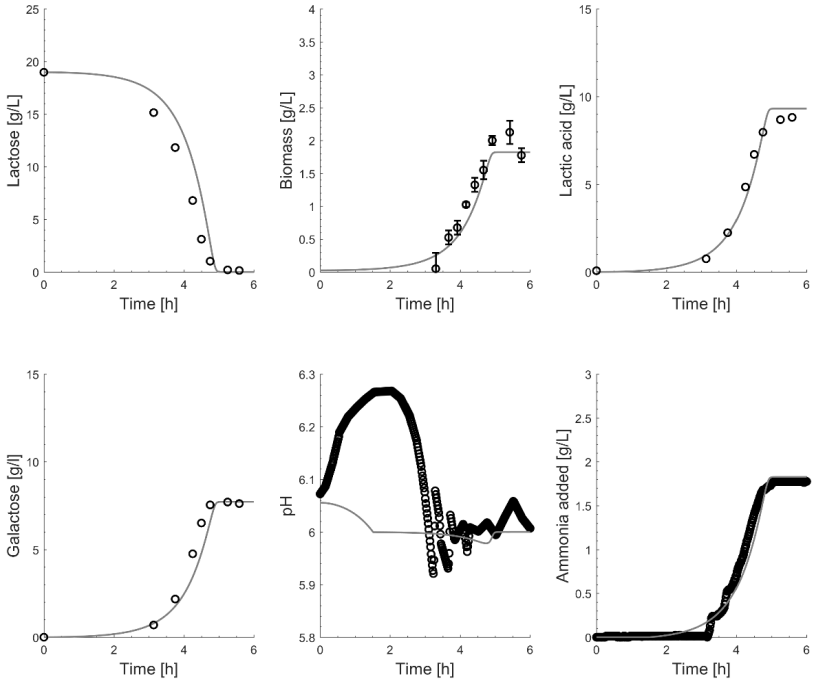


Fig. S5. Model prediction for a validation lab-scale batch fermentation. The *S. thermophilus* fermentation was performed in a 2 L stirred tank bioreactor at 300 rpm, 40 °C, and controlled at pH = 6. Model prediction (solid line) for the measurements (circles) of one of the two validation lab-scale batches. The biomass measurement is shown with the standard deviation.

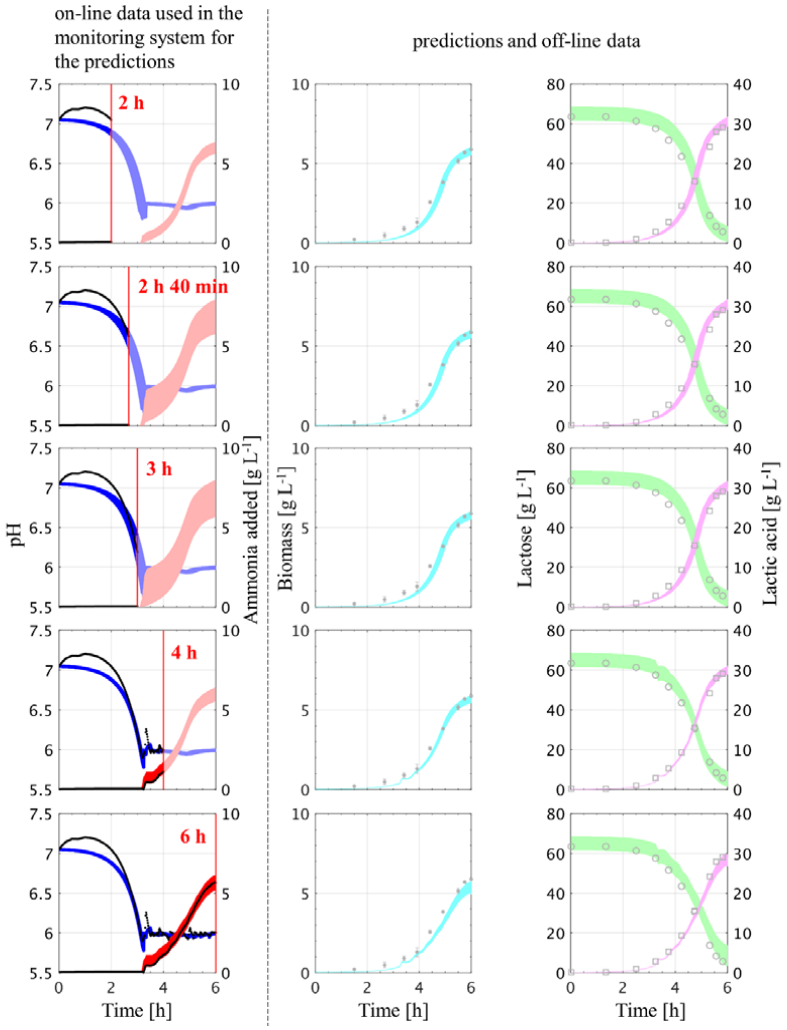


Fig. S6. Probabilistic monitoring system applied to lab-scale batch data of a *S. thermophilus* fermentation. The monitoring system reads in the on-line available data (black dots), ammonia addition and pH, and predicts the state variables every 5 minutes. 100 Monte Carlo simulations of the dynamic model were performed within the monitoring system considering uncertainties in the initial conditions, ammonia addition, and model parameters. The 95 % confidence intervals of the predictions are shown at five time points during the fermentation (2 h, 2 h 40 min, 3 h, 4 h, 6 h). Predictions of the pH (blue), ammonia addition (red), biomass (cyan), lactose (green), and lactic acid (magenta) concentrations. The off-line measurements for biomass (gray dot with standard deviation), lactose (gray circle), and lactic acid (gray square) are shown for comparison only, but were not used for the data reconciliation and parameter update (see Fig. 1).

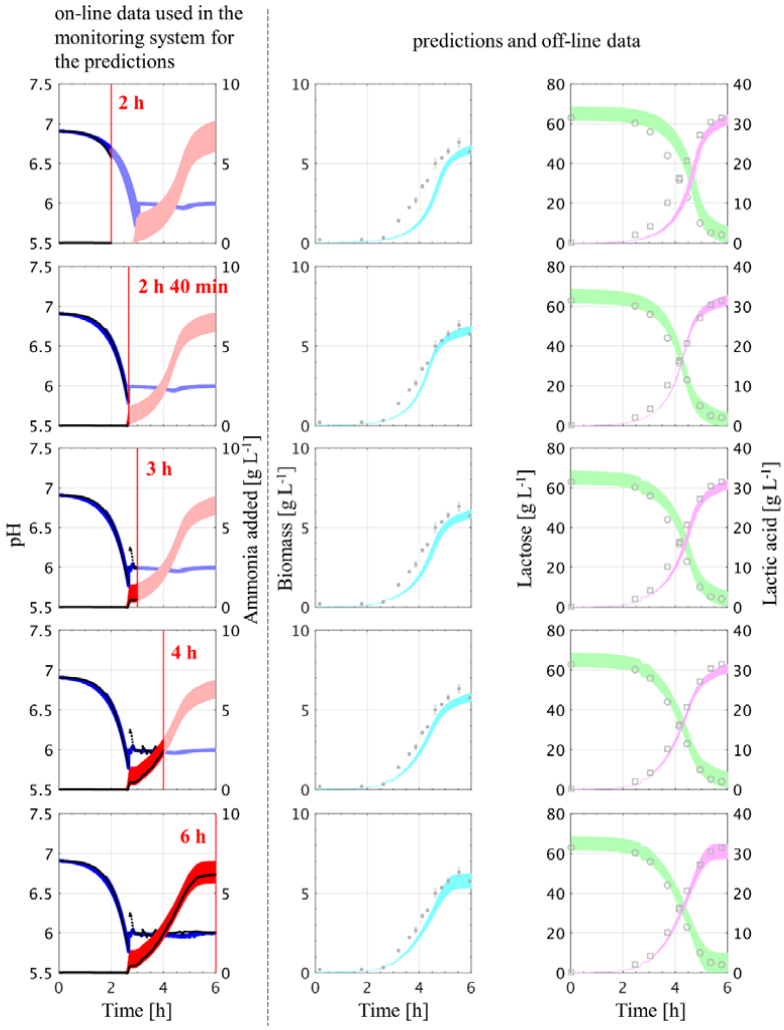


Fig. S7. Probabilistic monitoring system applied to lab-scale batch data of a *S. thermophilus* fermentation. The monitoring system reads in the on-line available data (black dots), ammonia addition and pH, and predicts the state variables every 5 minutes. 100 Monte Carlo simulations of the dynamic model were performed within the monitoring system considering uncertainties in the initial conditions, ammonia addition, and model parameters. The 95 % confidence intervals of the predictions are shown at five time points during the fermentation (2 h, 2 h 40 min, 3 h, 4 h, 6 h). Predictions of the pH (blue), ammonia addition (red), biomass (cyan), lactose (green), and lactic acid (magenta) concentrations. The off-line measurements for biomass (gray dot with standard deviation), lactose (gray circle), and lactic acid (gray square) are shown for comparison only, but were not used for the data reconciliation and parameter update (see Fig. 1).

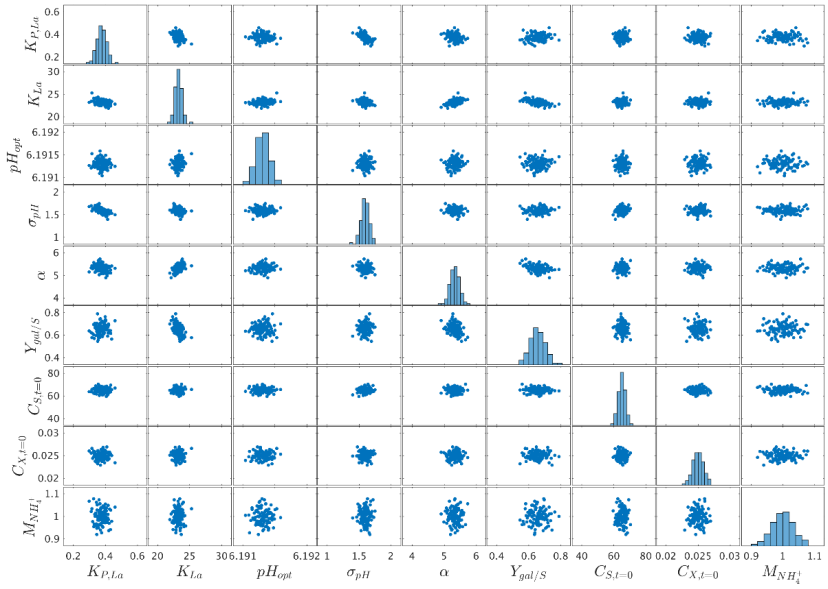


Fig. S8. Plotting of the sampling matrix for the input uncertainty space. The Latin Hypercube Sampling (LHS) technique and the Iman Conover rank correlation method were used to sample 100 independent inputs and to induce the known covariance matrix, respectively.

Stoichiometric matrix

	Stoichiometric matrix															
Components -> i																
Name	Lactose	Galactose	Biomass	Lactic acid	Lactate	Ammonia	Ammonium	Phosphoric acid	Dihydrogen phosphate	Hydrogen phosphate	Carbon dioxide dissolved = carbonic acid	Bicarbonate	Hydrogen ion	Hydroxyl ion	unknown undissociated form acid	unknown dissociated from acid
Symbol	**S**	**Gal**	**X**	**HLa**	**La1-**	**A**	**A1-**	**P**	**P1-**	**P2-**	**Cl**	**C1-**	**H**	**OH**	**Z+**	**Z**
Chemical composition	C12H22O11	C6H12O6	CHaObNcPd	C3H6O3	C3H5O3(-)	NH3	NH4(+)	H3PO4	H2PO4(-)	HPO4(2-)	H2CO3*(l)	HCO3(-)	H+	OH-		
Process (Units)	C-mol/L	C-mol/L	C-mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L	mol/L
Biomass growth	-(1+Ygal)	Ygal	1			c		d								
Lactic acid synthesis	-(1+Ygal)	Ygal			1/3								1/3			
Ammonia dissociation						1	-1						1			
Phosphate dissociation 1								-1	1				1			
Phosphate dissociation 2									-1	1			1			
Carbonate dissociation 1											-1	1	1			
Lactate dissociation				-1	1								1			
Water dissociation													1	1		
Dissociation of Z													1		-1	1

Paper C

A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model

Robert Spann[a], Christophe Roca[b], David Kold[b], Anna Eliasson Lantz[c], Krist V. Gernaey[a], Gürkan Sin[a]

[a] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] Chr. Hansen A/S

[c] PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark

Published in the Proceedings of the 27th European Symposium on Computer Aided Process Engineering – ESCAPE 27, 2017.
http://dx.doi.org/10.1016/B978-0-444-63965-3.50372-X

A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model

Robert Spann[a], Christophe Roca[b], David Kold[b], Anna Eliasson Lantz[a], Krist V. Gernaey[a] and Gürkan Sin[a*]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Building 229, 2800 Kgs. Lyngby, Denmark*
[b]*Chr. Hansen, Boege Allé 10-12, 2970 Hoersholm, Denmark*
gsi@kt.dtu.dk

Abstract

Lactic acid bacteria are used in many industrial applications, e.g. as starter cultures in the dairy industry or as probiotics, and research on their cell production is highly required. A first principles kinetic model was developed to describe and understand the biological, physical, and chemical mechanisms in a lactic acid bacteria fermentation. We present here a consistent approach for a methodology based parameter estimation for a lactic acid fermentation. In the beginning, just an initial knowledge based guess of parameters was available and an initial parameter estimation of the complete set of parameters was performed in order to get a good model fit to the data. However, not all parameters are identifiable with the given data set and model structure. Sensitivity, identifiability, and uncertainty analysis were completed and a relevant identifiable subset of parameters was determined for a new parameter estimation including an evaluation of the correlation and confidence intervals of those parameters to double-check identifiability issues. Such a consistent approach supports process modelling and understanding as i.e., one avoids questionable interpretations caused by estimates of actually unidentifiable parameters.

Keywords: lactic acid bacteria, parameter estimation, sensitivity analysis, identifiability analysis, uncertainty analysis

1. Introduction

Biotechnological manufacturing companies increasingly apply first principles models in order to support the development, optimization, and control of their processes. Models are an indispensable tool for the comprehensive understanding and control of the processes (Mears et al., 2016). Within the model development process, model parameters need to be estimated. However, often an ill-conditioned parameter estimation is performed because no statistical evaluation of the procedure is conducted and the selected set of estimated parameters is not identifiable.

In this work, a consistent statistical parameter estimation methodology is followed for an unstructured first principles model of a lactic acid bacteria fermentation. This model can be applied in a process monitoring system that allows risk-based operation and decision-making. The aim of this study is to accomplish a consistent statistical analysis of the estimated parameters. To this end, a local sensitivity and identifiability analysis are applied.

2. Nomenclature

Symbol	Description		
C_i	concentration (g L^{-1})	α	growth related production coefficient of
K_I	substrate inhibition constant (g L^{-1})		lactic acid (g g^{-1})
K_S	limiting substrate constant (g L^{-1})	β	non-growth related production coef. of
K_{La}	lactate inhibition parameter (L g^{-1})		lactic acid (g g^{-1} h^{-1})
K_{HLa}	lactate inhibition parameter (g L^{-1})	σ	std. deviation in the pH function
$K_{P,HLa}$	2. HLa inhibition parameter (L g^{-1})		
P	lactic acid (g L^{-1})	Subscripts i	
pH_{opt}	optimal pH in the pH function	HLa	undissociated lactic acid
t_{lag}	lag-time coefficient (h)	La	dissociated lactic acid
X	biomass (g L^{-1})	P	product: lactic acid
μ_{max}	max. specific growth rate (h^{-1})	S	substrate: lactose

3. Materials and Methods

3.1. Fermentation

The batch fermentation of *Streptococcus thermophilus* was carried out in a 2 L stirred tank glass bioreactor (Biostat® B, Sartorius AG, Germany) with a stirring speed of 300 rpm and nitrogen headspace gassing. The temperature was maintained at 40 °C and the pH was controlled at pH = 6 by the addition of 24 % (v/v) NH_3. The fermentation medium contained 18 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$.

3.2. Analysis

Dry cell weight was determined from cell broth, which was centrifuged, washed in 0.9 % (w/v) NaCl solution, and dried for 24 h at 70 °C. Organic acids were quantified from filtered (pore size: 0.2 µm) supernatant samples in a HPLC system (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA) and a refractive index detector (ERC RefractoMax 520), equipped with an Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA) using 5 mM H_2SO_4 at a flow rate of 0.6 mL min^{-1} at 50 °C.

3.3. Simulation and Modelling

An unstructured first principles model describing the homolactic batch fermentation of *Streptococcus thermophilus* was developed. Lactose is metabolized yielding biomass, lactic acid, and galactose. The present strain consumes the glucose part from lactose completely, while galactose is mostly secreted. The growth rate expression in Eq. (1) depends on the lag time, lactose inhibition and limitation (Åkerberg et al., 1998), lactic acid inhibition, which is separated into the dissociated and undissociated form (Aghababaie et al., 2015), and the pH of the cultivation (Schepers et al., 2002). An amended form of the Luedeking-Piret (Luedeking and Piret, 1959) equation describes the lactic acid production in Eq. (2).

$$\frac{dX}{dt} = \mu_{max} \cdot f_{lag} \cdot f_S \cdot f_{La} \cdot f_{HLa} \cdot f_{pH} \cdot X = \mu_{max} \cdot (1 - e^{-t/t_{lag}}) \cdot \frac{C_S}{C_S + K_S + \frac{C_S^2}{K_I}} \cdot e^{-K_{La} \cdot C_{La}} \cdot \frac{1}{1 + e^{K_{P,HLa}(C_{HLa} - K_{HLa})}} \cdot e^{-\left(\frac{(|pH_{opt} - pH|)^2}{2\sigma^2}\right)} \cdot X \quad (1)$$

$$\frac{dP}{dT} = (\alpha \cdot \mu + \beta \cdot f_S \cdot f_{pH}) \cdot X \quad (2)$$

The kinetic model has been implemented in MATLAB® (The MathWorks®, Natick, MA) and the ODEs were solved with the ode15s solver.

3.4. Consistent Statistical Parameter Estimation

The consistent statistical parameter estimation follows the methodology from Figure 1. In the beginning, just literature values or expert knowledge are available for the initial parameter guess, but they do often not refer to the exact strain or cultivation conditions of interest. The first parameter estimation of the complete set of parameters is performed with this weak initial parameter guess. In the present study, the statistical parameter estimation and the uncertainty analysis with the Monte Carlo method were solved in MATLAB applying the methodology of Sin and Gernaey (2016). In the statistical assessment, the parameters are first ranked according to their significance on the model outputs, measured by δ^{msqr}, according to the methodology of Brun et al. (2002). The least squares method is applied for the sensitivity and identifiability analysis as defined in Seber and Wild (1989). Then, the confidence intervals are checked. Third, the covariance matrix is evaluated and parameters with a pairwise correlation larger than 0.5 are regarded as correlated. Parameters with too large confidence intervals and with zero influence on the model outputs are not identifiable and excluded. A new parameter estimation is performed with a parameter subset with a collinearity index below 15. It is also advised to run the parameter estimation of several subsets, if there is no obvious subset candidate, and then to compare the results. The results of the new parameter estimation are again statistically evaluated. The parameter estimate is considered as identifiable when the pairwise correlation of the parameters is lower than 0.5 and the relative error (RE = standard deviation / mean value of the parameter value) is lower than 0.1. Otherwise, the parameter estimation and statistical evaluation have to be repeated until a statistically identifiable parameter subset is obtained or a new design of experiments has to be performed in order to obtain more experimental information.

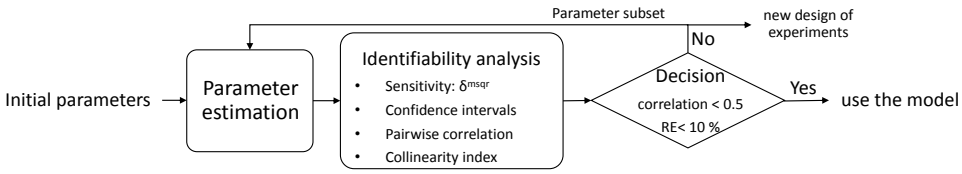


Figure 1: Steps for the consistent statistical parameter estimation.

Table 1: First parameter estimation: initial values and results.

kinetic parameters	initial values	source of initial values	estimated parameter values	95% CI	δ^{msqr} lactose
μ_{max}	1.18 h^{-1}	(Aghababaie et al., 2015)	2.89	±9	12.8
K_S	1 g L^{-1}	expert guess	0.5	±6	0.5
K_I	164 g L^{-1}	(Åkerberg et al., 1998)	47	±683	3.7
K_{La}	0.326 L g^{-1}	(Aghababaie et al., 2015)	3E-7	±0.3	0
pH_{opt}	6	expert knowledge	6	±15,700	0
σ	1.54	(Schepers et al., 2002)	5	±19,653	0
α	1.54 g g^{-1}	(Aghababaie et al., 2015)	6E-3	±11	0
β	0.52 g g^{-1} h^{-1}	(Aghababaie et al., 2015)	7.54	±32	3.3
t_{lag}	1 h	expert guess	0.84	±5	3.7

4. Results and Discussion

The consistent statistical parameter estimation was applied to the fermentation data set, as described in the Materials and Methods section. The model parameters were estimated in order to fit all four offline measurements during the entire fermentation: biomass, lactose, lactic acid, and galactose. Initial parameter values were obtained from literature and fermentation specialists (Table 1). However, one has to be aware of that both the strain is not identical and the model structure was slightly changed compared to the literature. For example, the maximum specific growth rate parameter μ_{max} did therefore not reflect the biological specific maximum growth rate directly since it has to compensate for all functions in the growth rate expression, and should be better interpreted as an artificial parameter. The undissociated lactic acid inhibition parameter K_{HLa} = 0.0444 g L^{-1} and the second lactate inhibition parameter $K_{P,HLa}$ = 52.862 L g^{-1} given from Aghababaie et al. (2015) were kept constant as they could not be estimated with the given experimental data set because the concentration of the undissociated form was too low at the given pH = 6. Table 1 summarizes the results of the first parameter estimation, which was within the expected range. Figure 2 shows the initial model fit as dashed line. Considering that μ_{max} accounts for five functions in the growth rate expression, a biological μ_{max} would be in the range of 1.7 h^{-1} when assuming that each function has to be compensated by 10 % on average. The model parameters were ranked by their local significance using the delta mean-square measure,

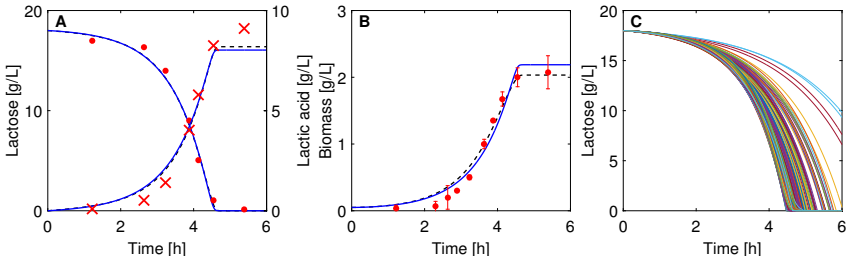


Figure 2: *Streptococcus thermophilus* batch fermentation. A and B: The offline data lactose and biomass (dots), lactic acid (crosses), the initial model fit (dashed line), and the second model fit (line). C: Monte Carlo simulation (N=1000) with the uncertainty of the parameter estimates for the model output lactose.

Table 2: Correlation matrix of the 9 parameters in the first parameter estimation.

	μ_{max}	K_S	K_I	K_{La}	pH_{opt}	σ	α	β	t_{lag}
μ_{max}	1	0.72	-0.67	-0.15	-0.72	0.62	0.25	0.67	-0.57
K_S		1	-0.69	-0.55	-0.43	0.32	-0.08	0.81	-0.7
K_I			1	0.14	0.04	0.04	-0.58	-0.25	0.95
K_{La}				1	0.34	-0.24	0.16	-0.49	0.33
pH_{opt}					1	-0.9	0.18	-0.75	0.03
σ						1	-0.17	0.64	0.07
α							1	-0.49	-0.48
β								1	-0.29
t_{lag}									1

δ^{msqr}. δ^{msqr} values of the parameters with respect to lactose are listed in Table 1. K_{La}, pH_{opt}, σ, and α had zero effect on the model outputs. The large confidence intervals showed that K_I, pH_{opt}, and σ could not be identified with the given data set. The data set lacked information of different substrate concentrations and pH conditions in order to determine them. Table 2 shows the correlation matrix of the 9 investigated parameters. The correlation coefficient was higher than 0.5 for several parameters indicating high linear correlations between these model parameters. For example, μ_{max} was highly correlated with K_S and K_I as well as with the parameters from the pH and lag time functions. α and β had a correlation coefficient of 0.49, which did not clearly show correlation. The explanation was that the estimated value of α was very low. K_I, pH_{opt}, σ, K_{La} and α were excluded from the next parameter estimation based on the sensitivity analysis and the large confidence intervals. It was hence decided to set the functions f_{La}, f_{HLa}, and f_{pH} equal to one. α was kept at the initial literature value and β was estimated. The collinearity index for this subset of parameters was 4.6, indicating identifiability (data not shown). It was noted that other subsets with more parameters and a collinearity index below 15 existed, but they included also parameters that had no effect on the model output. The second parameter estimation for the selected four parameters gave the same quality of fit as shown in Figure 2. The new estimated parameter values are listed in Table 3 including their confidence intervals, δ^{msqr} with respect to lactose, and the correlation matrix. The parameters were still in the range known from literature. They had furthermore all a significant effect on the model outputs. The relative errors were however large for K_S and t_{lag} indicating a bad quality of the parameter estimation (data not shown). The parameters were still correlated. The correlation coefficients differed from the previous matrix because another parameter subset was investigated. In order to predict the uncertainty, i.e. the effect of the estimated parameters and their standard deviation on the model outputs, a Monte-Carlo simulation was performed. Figure 2 indicates that the uncertainties in the parameter estimates had a substantial effect on the model outputs.

This study verified the methodology for a consistent statistical parameter estimation. However, the problem is not completely solved for the given data set and model structure because the data set does not contain enough information in order to estimate the parameters accurately, and the model needs to be simplified. To improve the parameter estimation accuracy, it is necessary to perform a new design of experiments, e.g. with different initial conditions so as to obtain information rich new data for parameter estimation.

Table 3: Results of the second parameter estimation with the confidence intervals (CI).

kinetic parameters	estimated parameter values	95% CI	δ^{msqr} lactose	correlation matrix			
				μ_{max}	K_S	β	t_{lag}
μ_{max}	1.97	$\pm$0.66	14.3	1	0.87	0.64	0.97
K_S	0.53	$\pm$1.24	0.7		1	0.55	0.76
β	4.2	$\pm$1.8	2.2			1	0.76
t_{lag}	0.81	$\pm$0.94	3.7				1

5. Conclusion

First principles models are increasingly applied in biotechnological manufacturing processes and reliable parameter estimation procedures are needed. This study verified the methodology for a consistent statistical parameter estimation. However, the given problem was not solved completely since the available data set did not include enough information to estimate all parameters accurately. The estimated parameters contained correlation. It is therefore necessary to acquire more fermentation data, which also covers different fermentation conditions. In order to identify for example K_S more measurement points at low lactose concentrations are needed. Furthermore, the kinetic model is too complex. A simplification of it is recommended.

6. Acknowledgement

Chr. Hansen is thanked for the experimental support. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 643056.

References

Aghababaie, M., Khanahmadi, M., Beheshti, M., 2015. Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures. Food and Bioproducts Processing 94, 657–667.

Åkerberg, C., Hofvendahl, K., Hahn-Hägerdal, B., Zacchi, G., 1998. Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour. Applied Microbiology and Biotechnology 49, 682–690.

Brun, R., Kühni, M., Siegrist, H., Gujer, W., Reichert, P., 2002. Practical identifiability of ASM2d parameters - Systematic selection and tuning of parameter subsets. Water Research 36, 4113–4127.

Luedeking, R., Piret, E. L., 1959. Kinetic study of the lactic acid fermentation. Batch process at controlled pH. Biotechnology and Bioengineering 67, 636–644.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K. V., 2016. Mechanistic Models for Process Development and Optimization of Fed-batch Fermentation Systems. In: Kravanja, Z. (Ed.), Proceedings of the 26th European Symposium on Computer Aided Process Engineering – ESCAPE 26.

Schepers, W. A., Thibault, J., Lacroix, C., 2002. Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. Enzyme and Microbial Technology 30, 187–194.

Seber, G. A. F., Wild, C. J., 1989. Nonlinear Regression, 792.

Sin, G., Gernaey, K. V., 2016. Data Handling and Parameter Estimation. In: Experimental Methods in Waste-water Treatment. pp. 201–234.

Paper D

Model-based Process Development for a Continuous Lactic Acid Bacteria Fermentation

Robert Spann[a], Anna Eliasson Lantz[b], Christophe Roca[c], Krist V. Gernaey[a], Gürkan Sin[a]

[a] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark

[c] Chr. Hansen A/S

Published in the Proceedings of the 28th European Symposium on Computer Aided Process Engineering – ESCAPE 28, 2018.
https://doi.org/10.1016/B978-0-444-64235-6.50279-5

Model-based process development for a continuous lactic acid bacteria fermentation

Robert Spann[a], Anna Eliasson Lantz[b], Christophe Roca[c], Krist V. Gernaey[a] and Gürkan Sin[a*]

[a]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Building 229, 2800 Kgs. Lyngby, Denmark*
[b]*PILOT PLANT, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads Building 229, 2800 Kgs. Lyngby, Denmark*
[c]*Chr. Hansen A/S, Bøge Allé 10-12, 2970 Hørsholm, Denmark*
gsi@kt.dtu.dk

Abstract
A mechanistic process model describing a lactic acid bacteria (LAB) fermentation was applied to develop a continuous fermentation process. Producing LAB for the dairy industry in a continuous cultivation, which would allow harvesting the cells during the cultivation, would reduce production costs compared to traditional batch processes. To this end, a validated mechanistic model of a *Streptococcus thermophilus* fermentation was used for a model-based continuous process evaluation. The fermentation model consists of biological and chemical mechanisms including a description of the growth rate as a function of pH and inhibition effects of metabolites. The optimal dilution rate and substrate concentration in the feed were estimated in order to maximize the cell yield (biomass concentration) and to minimize the waste of substrate during the continuous fermentation in a 50 m^3 bioreactor for two scenarios: downstream capabilities are i) flexible, and ii) fixed. The biomass concentration is restricted by the growth-inhibiting lactic acid concentration, which is produced by the growing bacteria. Furthermore, the substrate, which is supplied by the feed, should be consumed completely in the fermentation and not wasted in the bioreactor effluent owing to raw material costs. The resulting non-linear optimization problem was formulated and solved in MATLAB®. A Monte Carlo simulation showed the robustness of the results, where a biomass concentration of 5 g L^{-1} could be achieved in the continuous fermentation with a substrate wastage of less than 3 % in the bioreactor effluent. The productivity of the continuous process was similar to a traditional batch process, but frequent cleaning and sterilization are no longer necessary in a continuous process resulting in a shorter unproductive downtime of the bioreactors. This promising potential of a continuous process for LAB cultivations encourages pilot-scale studies for a comprehensive techno-economic evaluation.

Keywords: continuous lactic acid bacteria fermentation, modelling, process development, Monte Carlo simulation

1. Introduction
Mechanistic models are increasingly applied in the biotechnological industry in order to develop, understand, optimize, monitor, and control fermentation processes. They can support the development of fermentation processes and give insight into the process during operation. There are several applications for mechanistic models: E.g. they may be applied offline to determine appro-

priate process operation conditions and control strategies, or online as soft sensor for model-based monitoring and control (Mears et al., 2017).

In this study, a mechanistic model describing a *Streptococcus thermophilus* fermentation was applied to design a continuous fermentation process. The food industry strives for continuous fermentations to reduce production costs compared to traditional batch fermentations, which are still the standard operation. The model predicted biological state variables, such as the biomass, substrate (lactose), and lactic acid concentrations. In addition, the model predicted the pH of the fermentation broth by solving the dissociation reactions of the charged components, as lactate, ammonia, carbonate and phosphate. The aim of this study was to propose the optimal process conditions for a continuous lactic acid fermentation to design a beneficial process compared to a batch process. To this end, the model was utilized to optimize the dilution rate and substrate concentration in the feed stream by maximizing the biomass concentration and reducing the substrate waste in the bioreactor effluent.

2. Nomenclature

Symbol	Description		
C_i	concentration (g L^{-1})	μ_{max}	max. specific growth rate (h^{-1})
F	medium flow rate (L h^{-1})	α	growth related production coefficient of lactic acid (g g^{-1})
K_I	substrate inhibition constant (g L^{-1})		
K_S	limiting substrate constant (g L^{-1})	σ_{pH}	std. deviation in the pH function
K_{La}	lactate inhibition parameter (L g^{-1})		
$K_{P,La}$	2. lactate inhibition param. (L g^{-1})		
$K_{P,pH1}$	lactate inhibition pH parameter	Subscripts i	
$K_{P,pH2}$	2. lactate inhibition pH parameter	La	dissociated lactic acid
pH_{opt}	optimal pH in the pH function	P	by-product: lactic acid
t_{lag}	lag-time coefficient (h)	S	substrate: lactose
V	volume of the bioreactor (L)	S_i	lactose in the feed inlet
Y_{gal}	galactose yield (g g^{-1})	X	biomass

3. Materials and Methods

3.1. Fermentation Model

A first principles model describing the homolactic *S. thermophilus* fermentation was applied for a continuous fermentation. The model was validated previously with 2 L batch fermentations at different pH set points (pH 5.5 - 7.0) and initial lactose (substrate) concentrations (20 and 70 g L^{-1}). Furthermore, the model was validated with a continuous accelerostat fermentation at lab-scale to ensure that the model predicts limiting substrate conditions as they occur in continuous cultivations, as well (manuscript in preparation). The studied strain consumes lactose, yielding biomass, lactic acid, and galactose because galactose is not metabolized but secreted. In the present study, the bacterial cells (biomass) are the product of interest. The growth rate expression (Eq. (1)) depends on the lag time, lactose inhibition and limitation (Åkerberg et al., 1998), pH dependent lactate inhibition, and the pH of the cultivation broth (Schepers et al., 2002).

$$\frac{dC_X}{dt} = \mu_{max} \cdot (1 - e^{-t/t_{lag}}) \cdot \frac{C_S}{C_S + K_S + \frac{C_S^2}{K_I}} \cdot \frac{1}{1 + e^{K_{P,La} \cdot (C_{La} - K_{La} \cdot \frac{1}{1 + e^{K_{P,pH1} * (pH - K_{P,pH2})}})}} \cdot e^{-\left(\frac{(pH_{opt} - pH)^2}{\sigma_{pH}^2}\right)} \cdot C_X - \frac{F}{V} \cdot C_X \quad (1)$$

Table 1: Kinetic parameters for the dynamic model of the *S. thermophilus* fermentation.

Parameter	Unit	Value	95 % confidence interval (CI)		Reference
K_I	g L^{-1}	164			(Åkerberg et al., 1998)
K_{La}	g L^{-1}	19.80	19.71	19.89	parameter estimation
$K_{P,La}$	L g^{-1}	0.24	0.17	0.30	parameter estimation
$K_{P,pH1}$		20			expert knowledge*
$K_{P,pH2}$		7			expert knowledge*
K_S	g L^{-1}	0.79			(Åkerberg et al., 1998)
pH_{opt}		6.38	6.28	6.50	parameter estimation
t_{lag}	h	1			expert knowledge*
Y_{gal}	g g^{-1}	0.69	0.61	0.76	parameter estimation
α	g g^{-1}	5.19	5.18	5.20	parameter estimation
μ_{max}	h^{-1}	2.06	2.03	2.08	parameter estimation
σ_{pH}	h^{-1}	1.42	1.35	2.50	parameter estimation

*) consultation of process experts.

The expression for the lactic acid secretion rate as by-product (Eq. (2)) is considered to be growth dependent (Peng et al., 1997).

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} - \frac{F}{V} \cdot C_P \tag{2}$$

The lactose uptake rate expression (Eq. (3)) is the sum of the growth and lactic acid secretion rate expressions taking the secretion of galactose (Y_{gal}) into account.

$$\frac{dC_S}{dt} = \frac{F}{V} \cdot (C_{S_i} - C_S) - (1 + Y_{gal}) \cdot (\frac{dC_X}{dt} + \frac{dC_P}{dt}) \tag{3}$$

The chemical model (for the pH calculation) comprises the dissociation reactions of the charged components in the fermentation broth, such as lactate, ammonia, carbonate, phosphate, and water (Musvoto et al., 2000). The kinetic model has been implemented in MATLAB® (The MathWorks®, Natick, MA) and the ODEs were solved with the ode15s solver. The kinetic parameters (Table 1) were derived from a consistent parameter estimation as described in Spann et al. (2017).

3.2. Design of the continuous fermentation conditions

The continuous process was a single stage chemostat cultivation without recirculation, hence the process is continuously fed with the growth medium and the culture broth is withdrawn at the

Table 2: Description of the design problem (LB: lower bound, UB: upper bound).

Scenario	Description	Flow rate [m^3 h^{-1}]		C_{Si} [g L^{-1}]	
		LB	UB	LB	UB
Scenario 1	50 m^3 bioreactor downstream flexible	1	20	50	150
Scenario 2	50 m^3 bioreactor downstream restricted	fixed to 10		50	150

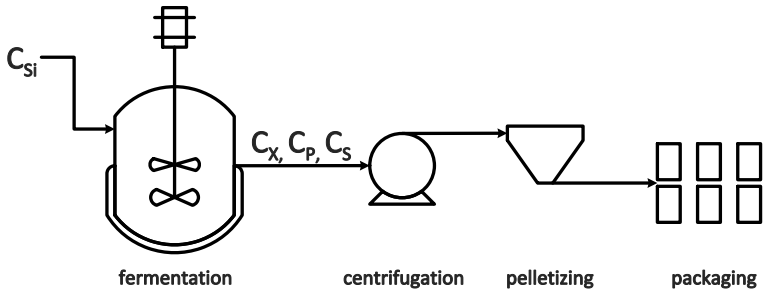


Figure 1: Scheme of the continuous lactic acid bacteria fermentation and downstream units.

same rate in order to maintain the bioreactor volume constant. The target product of the fermentation process are the cells, as they are used in the dairy industry for yoghurt and cheese production, subsequently. Optimal process conditions for a continuous *S. thermophilus* fermentation process were predicted for two scenarios (Table 2): Scenario 1) A 50 m^3 bioreactor is available and the downstream units will be built based on the calculated bioreactor effluent flow rate (Fig. 1). Scenario 2) A 50 m^3 bioreactor and the downstream units for centrifugation and pelletizing are already available with given capacities. The flow rate is fixed to 10 m^3 h^{-1} here. For scenario 1, the feed flow rate and lactose concentration in the feed solution were optimized. For scenario 2, only the lactose concentration in the feed solution was optimized. The carbon source lactose will be growth limiting. The challenge of the design problem for an optimal feed flow rate and inlet substrate concentration was that the lactic acid concentration in the broth is determined by the biomass concentration but lactic acid inhibits biomass growth at the same time. The objective was: i) to maximize the biomass concentration in the bioreactor effluent; and, ii) to minimize the waste of substrate in the bioreactor effluent for both scenarios. The objective function was solved in MATLAB using the built in solver, fmincon. The mean values of the biological model parameters were used in the optimization step.

3.3. Monte Carlo simulation

100 independent input samples were created in the probability range [0 1] with the Latin Hypercube Sampling (LHS) technique (McKay et al., 1979; Sin et al., 2009). Then, the Iman Conover method (Iman and Conover, 1982) was applied to induce the correlation between the model parameters, which was derived from a parameter estimation. The samples were finally converted into the real parameter space using the Matlab function icdf, in which all uncertainties were assumed to be normally distributed. A Monte Carlo simulation of the continuous fermentation process was performed considering both the biological model parameter uncertainties (Table 1) and variations of the lactose concentration in the feed (which was obtained in the previous optimization step) with $3\sigma = 10$ %.

4. Results and Discussion

Two scenarios were investigated considering both the new construction of the production facilities (scenario 1), and the limitations of an already available plant (scenario 2). For scenario 1, the capacities of the downstream units were considered to be flexible. A flow rate of 5 m^3 h^{-1} was estimated with a lactose concentration of 54 g L^{-1} in the feed solution (Fig 2). For scenario 2, an already existing production plant was assumed given limitations in the flow rates. Here, the estimated lactose concentration in the feed solution was 50 g L^{-1}. More lactose would be wasted in the bioreactor effluent in scenario 2 since the dilution rate is almost twice as high as in scenario 1, and the lactic acid bacteria cannot consume all the supplied lactose at this high rate.

The continuous fermentation starts with a batch phase with 65 g L^{-1} lactose initially (Fig. 3). During the exponential growth in the batch phase, the biomass concentration increased before the continuous fermentation started. The optimal time to switch from the batch to continuous mode will not be easily detectable since there is no dissolved oxygen (DO) tension signal in this non-aerated cultivation, where DO is usually used in aerated cultivations to monitor biomass activity. However, the base addition rate will in this case indicate when growth slows down as lactic acid secretion is diminished at the same time (data not shown). The Monto Carlo simulations considering model parameter uncertainties and variations of the lactose concentration in the feed inlet show that the considered uncertainties had a minor effect on the target biomass concentration with acceptable $3\sigma < 0.5$ g L^{-1}.

A continuous cultivation could be a profitable alternative to a batch process for the industrial production of lactic acid bacteria as the cells are the desired product that can be produced in a continuous manner, similar to primary metabolites. In addition, naturally derived lactic acid bacteria might be less prone to genetic instability, which might for example be an issue in a continuous cultivation with genetically modified organisms that are often used in the pharmaceutical industry. However, the application of a continuous cultivation is quite limited in the biotechnological industry so far. This might be mostly because of the risk for process contaminations, which could cause more harm in a continuously operated reactor than a batch culture. Nevertheless, especially in the presented lactic acid bacteria cultivation, the lactic acid concentration is already inhibiting the growth of many contaminants as the lactic acid concentration will be higher than 25 g L^{-1}. In a traditional batch process with an initial lactose concentration of 65 g L^{-1} a biomass concentration of 6 g L^{-1} was reached (see the initial batch phase in Fig. 3). The batch process will be, however, more expensive than the continuous culture due to more frequent cleaning, sterilization, pre-culture preparation tasks, etc., provided that the continuous process could be maintained stable with a low risk of failure.

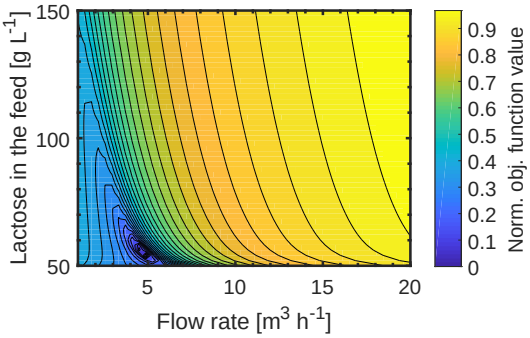


Figure 2: Contour plot of the normalized objective function value with respect to the lactose concentration in the feed and the feed flow rate. The minimum for scenario 1 is at F = 5 m^3 h^{-1} and C_{Si} = 54 g L^{-1}. The minimum for scenario 2 is at the lower limit C_{Si} = 50 g L^{-1} with the given F = 10 m^3 h^{-1}.

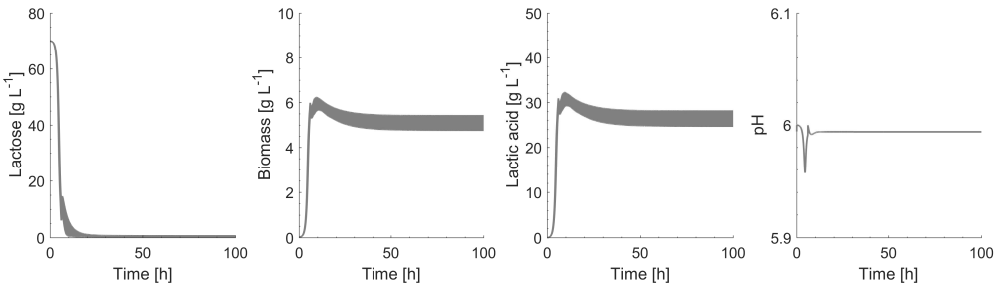


Figure 3: A continuous *S. thermophilus* fermentation with a batch phase in the beginning. 95 % confidence bounds for the limiting substrate lactose, dry cell weight, lactic acid concentration, and the pH (from left to right). A Monte Carlo simulation of 100 input samples was performed considering model parameter uncertainties, and variations of the lactose concentration in the feed solution.

Mechanistic models, as applied in the present study, are preferably used for off-line cultivation process development because of their many advantages (Mears et al., 2017): Mechanistic models are often able to extrapolate to process conditions outside the conditions that were used to develop the model. In addition, mechanistic models provide a flexible model structure that could be adjusted to several processes by adjusting the model parameters. Furthermore, some companies prefer to apply mechanistic models, e.g. for control purposes, because they want to understand where the predictions derive from, instead of following a black-box prediction. Due to these advantages, one should accept a longer model development and validation time of a mechanistic model compared to black-box models.

Further research is necessary to validate the presented model predictions in lab- or pilot-scale experiments. Biotechnological companies and further studies could use the presented case study by also including an economical objective function, e.g. considering sales revenue and expenses, such as operation and raw material costs, in order to design the most profitable continuous cultivation.

5. Conclusion

This study used a mechanistic model describing the production of *S. thermophilus* in a fermentation process to propose the optimal conditions for a continuous cultivation. The fermentation conditions were estimated for two scenarios: downstream capabilities are i) flexible, and ii) fixed. A continuous process is regarded as more cost effective than a traditional batch process, and especially mechanistic models are valuable for process design purposes because they might be capable to predict also conditions which were not experimentally investigated beforehand. In conjunction with uncertainty considerations, a probability distribution of the model prediction was obtained and the uncertainty in the biomass prediction was low, which shows the robustness of the model for analysis. Further studies are needed to validate the model predictions experimentally at pilot-scale and economic issues could be included in the optimization problem.

6. Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project).

References

Åkerberg, C., Hofvendahl, K., Hahn-Hägerdal, B., Zacchi, G., 1998. Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour. Applied Microbiology and Biotechnology 49, 682–690.

Iman, R. L., Conover, W. J., 1982. A distribution-free approach to inducing rank correlation among input variables. Communications in Statistics - Simulation and Computation 11 (3), 311–334.

McKay, M. D., Beckman, R. J., Conover, W. J., 1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21 (2), 239–245.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K. V., 2017. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. Trends in Biotechnology 35 (10), 914–924.

Musvoto, E., Wentzel, M., Loewenthal, R., Ekama, G., 2000. Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems. Water Research 34 (6), 1857–1867.

Peng, R. Y., Yang, T. C. K., Wang, H.-e., Lin, Y.-c., Cheng, C., 1997. Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model. Journal of the Chinese Agricultural Chemical Society 35 (5), 485–494.

Schepers, W. A., Thibault, J., Lacroix, C., 2002. Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. Enzyme and Microbial Technology 30, 187–194.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M., Gujer, W., 2009. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. Water Research 43 (11), 2894–2906.

Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., Sin, G., 2017. A Consistent Methodology Based Parameter Estimation for a Lactic Acid Bacteria Fermentation Model. In: Espuña, A., Graelss, M., Puigjaner, L. (Eds.), Proceedings of the 27th European Symposium on Computer Aided Process Engineering - ESCAPE 27. Elsevier, p. 3042.

Paper E

CFD Predicted pH Gradients in Lactic Acid Bacteria Cultivations

Robert Spann[a], Jens Glibstrup[a], Klaus Pellicer Alborch[b], Stefan Junne[b], Peter Neubauer[b], Christophe Roca[c], David Kold[c], Anna Eliasson Lantz[a], Gürkan Sin[a], Krist V. Gernaey[a], Ulrich Krühne[a]

[a] Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] Department of Biotechnology, Technische Universität Berlin

[c] Chr. Hansen A/S

In preparation to be submitted to Biotechnology and Bioengineering (Online ISSN:1097-0290).

Abstract

The formation of pH gradients in a 700 L batch fermentation of *Streptococcus thermophilus* was studied using multi-position pH measurements and computational fluid dynamic (CFD) modelling. To this end, a dynamic, kinetic model of *S. thermophilus* and a pH correlation were integrated into a validated one-phase CFD model, and a dynamic CFD simulation was performed. First, the fluid dynamics of the CFD model were validated with NaOH tracer pulse mixing experiments. Mixing experiments and simulations were performed while multiple pH sensors, which were placed vertically at different locations in the bioreactor, captured the response. A mixing time of about 46 s to reach 95 % homogeneity was measured and predicted at an impeller speed of 242 rpm. The CFD simulation of the *S. thermophilus* fermentation captured the experimentally observed pH gradients between a pH of 5.9 and 6.3, which occurred during the exponential growth phase. A pH higher than 7 was predicted in the vicinity of the base solution inlet. Biomass growth, lactic acid production, and substrate consumption matched the experimental observations. Moreover, the biokinetic results obtained from the CFD simulation were similar to a single-compartment simulation, for which a homogeneous distribution of the pH was assumed. This indicates no influence of pH gradients on growth in the studied bioreactor. This study verified that the pH gradients during a fermentation in the pilot-scale bioreactor could be accurately predicted using a coupled simulation of a biokinetic and a CFD model. In order to support the understanding and optimization of industrial-scale processes, future biokinetic CFD studies need to assess multiple types of environmental gradients, like pH, substrate, and dissolved oxygen, especially at industrial scale.

Introduction

Heterogeneities of culture parameters like substrate concentrations, pH, and dissolved oxygen concentrations are regarded as mainly responsible for productivity loss in large-scale bioreactor cultivations. Transport limitations occur at large scale due to insufficient mixing, and cells are consequently exposed to fluctuating conditions. Non-limiting substrate concentrations in the range of 0.3 – 2 g L^{-1} are reported in feeding zones during fed-batch processes, while there are substrate-limited conditions further away from the feeding position (Bylund et al., 1998; Larsson et al., 1996). pH values might also be outside of physiological ranges next to acid or base addition points (Langheinrich and Nienow, 1999; Lara et al., 2006). Mixing times of large-scale bioreactors for microbial cultures exceed 100 s to reach 95 % homogeneity, and the circulation time of the cells, which is proportional to the mixing time, is consequently in the magnitude of 10 s and longer depending on the stirring conditions (Delvigne et al., 2006; Nagata, 1975). Cells might adapt to continuously changing environments while they move through the bioreactor. Biomass and product yield reduction are reported for several different strains and processes when a process is scaled up to large scale (Bylund et al., 1998; Enfors et al., 2001; George et al., 1998; Xu et al., 1999). This is most likely related to heterogeneities, because microorganisms are exposed to fluctuating environmental conditions at large scale, which might affect the metabolic activity. pH gradients have shown an influence on the transcriptional response and enzyme activity in bacteria, and may therefore lead to decreased biomass growth and product formation as shown in scale-down studies (Amanullah et al., 2001; Cortés et al., 2016; Onyeaka et al., 2003).

Computational fluid dynamic (CFD) modelling is capable of representing the fluid dynamic conditions in bioreactors. It was already applied for process optimization by improving the impeller configuration for an increased oxygen transfer rate (Yang et al., 2012; Zou et al., 2012). Moreover, biokinetic models are coupled with fluid dynamics to analyze environmental

gradients during fermentations (Schmalzriedt et al., 2003; Wang et al., 2015). Either compartment models can be built and coupled with a biokinetic model or a biokinetic model is directly integrated into a CFD model. Compartment models, which are based on the knowledge about the fluid dynamics in the bioreactor obtained from CFD models, reduce the number of spatial elements and decrease the computational demand (Vrábel et al., 2001). If biokinetic models are directly integrated into CFD, both the Euler-Euler approach (Bannari et al., 2012; Elqotbi et al., 2013) and the Euler-Lagrange approach combined with a population balance model (Haringa et al., 2016; Lapin et al., 2004; Lapin et al., 2006; Morchain et al., 2013) are commonly applied. The fluid is treated as a continuum in both approaches, but the biological phase is treated as a continuum in the Euler-Euler approach and as a discrete phase in the Euler-Lagrange approach. This allows to track single cells there. So far, studies have mainly been focused on substrate and oxygen gradients in aerobic nutrient-limited fed-batch processes. Furthermore, their relevance is questionable because many of the aforementioned works use CFD models that were not experimentally validated e.g. by comparing the model response to mixing experiment data. There is therefore a considerable lack of scientific literature focusing specifically on dynamic CFD simulations of biokinetic models integrated into validated CFD models with the intention to simulate the formation of pH gradients in pilot and large-scale bioreactors.

The objective of this study was to predict the pH gradients, which occur in a 700 L bioreactor during a *Streptococcus thermophilus* fermentation, by coupling CFD and kinetic modelling in a CFD simulation. This tool, which combines fluid dynamics and microbial kinetics, will be used to study pH heterogeneities at pilot scale. To this end, first a one-phase CFD model of a 700 L bioreactor for a *S. thermophilus* fermentation was set up. Tracer pulse experiments with a NaOH solution and multi-position pH monitoring validated the fluid dynamic model predictions of the bioreactor. Then a kinetic model describing the biomass growth, lactic acid synthesis, and lactose consumption of *S. thermophilus* was integrated into the validated CFD model to simulate a pH controlled batch cultivation. An algebraic equation was applied to calculate the pH value based on the lactic acid and ammonia concentrations.

Materials and Methods

Bioreactor geometry and settings

A stirred tank bioreactor (Chemap AG, Switzerland) equipped with three 6-blade Rushton turbines was used (Fig. 1) and filled to a liquid height of H_L = 1.92 m, corresponding to a volume of 700 L, for both the mixing time experiment and the fermentation. The stirrer speed was 242 rpm (P/V = 0.79 kW m^{-3}) for the mixing time determination and 132 rpm (P/V = 0.13 kW m^{-3}) for the fermentation. The stirrer speed was measured with testo 477 LED stroboscope (Testo SE & Co. KGaA, Germany). The Reynolds number (Re) was defined as:

$$Re = \frac{N \cdot D_i^2 \cdot \rho_{H_2 0}}{\mu_{H_2 0}} \quad (1)$$

where N represents the stirrer speed, D_i the impeller diameter, ρ the fluid density, and μ the dynamic viscosity of the fluid.

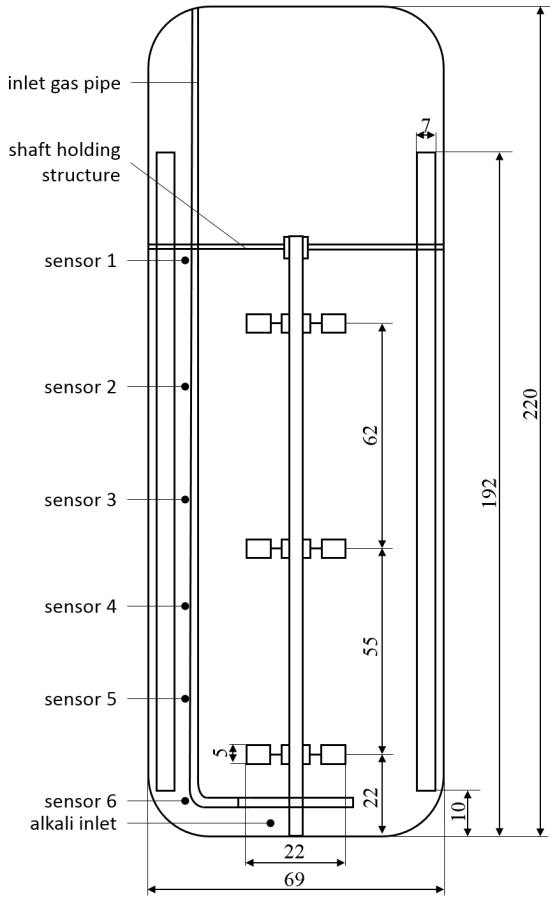


Figure 1. The geometry of the stirred tank bioreactor with dimensions in cm. The bioreactor is equipped with three 6-blade Rushton turbines and four baffles. Six monitoring points were positioned in the bioreactor. The reactor was filled up to 1.92 m liquid height both in the pulse experiment to determine the mixing time and in the fermentation.

The power input (P) was calculated:

$$P = N_P \cdot \rho \cdot N^3 \cdot D_i^5 \tag{2}$$

where N_P is the power number. N_P was assumed to be 5.5 (Doran, 1995; Ruston et al., 1950) for each Rushton turbine as $Re > 10^5$ (see the results section). The power input could unfortunately not be measured in the studied bioreactor.

Mesh generation and simulation settings

The bioreactor geometry was designed in SolidWorks (Dassault Systèmes, France). The sparger ring, the gas inlet pipe, and a supporting structure, which holds the shaft, were omitted. The bioreactor consisted of a stationary tank domain and three rotating impeller domains. Only half of the bioreactor volume was modelled applying a rotational periodicity plane. Meshes

with 6-sided hexahedral elements were defined for both domains in ANSYS ICEM CFD 17.1 (ANSYS, Inc., US-PA). The stationary domain consisted of approximately 2,000 mesh elements per liter. Each rotating domain, with a height of H_I = 0.075 m and diameter D_I = 0.32 m, was defined with about 95,000 elements per liter. The complete mesh consisted of 1.6 million nodes. The interface between the rotating and stationary domains was defined as Frozen-Rotor interfaces. CFD simulations were performed in ANSYS CFX 17.1 with the k-ε turbulence model (Supplementary Material). The top boundary was assumed a flat surface with a free-slip wall. The liquid density was assumed to be $\rho = 997\ kg\ m^{-3}$ and the dynamic viscosity $\mu = 8.9 \cdot 10^{-4}\ kg\ m^{-1}\ s^{-1}$ in both the tracer pulse and fermentation simulation.

Tracer pulse simulations

An additional variable was specified for the tracer pulse experiments in all domains with a diffusion coefficient $D_{tracer} = 5.17 \cdot 10^{-9}\ m^2\ s^{-1}$ corresponding to the characteristics of the tracer compound OH^- (Cents et al., 2005). Six monitoring points were located at different positions, which corresponded to the sensor positions (Fig. 1). Vertical positions of the sensors were, with respect to the bottom of the bioreactor: 0.10 m, 0.35 m, 0.60 m, 0.95 m, 1.25 m, and 1.60 m. The horizontal distance to the bioreactor wall was 0.10 m. A transient (time-dependent velocity field) simulation was performed with a physical time scale of a step time of 0.1 s and a root mean square (RMS) residual target of 10^{-4}. The RMS is a measure to validate the convergence. The tracer pulse was simulated by starting the simulation with 1 mol of the tracer variable in a cylindrical volume of height of 0.20 m and a width of 0.10 m, which was at the center at the top of the liquid phase.

Tracer pulse experiments

The bioreactor was filled with tap water. An NaOH solution (27 %, Novadan ApS, Denmark) was used as tracer substance for the pulse experiments. When dosing a pulse, 80 mL of NaOH was poured from 0.30 m above the liquid level at the center of the bioreactor into the liquid within 1 s. Experiments were carried out at 35 °C within a pH range of 5.0 to 6.0 to ensure that the mixing time is measured without interference of the slow reverse reaction of the carbonate dissociation (Einsele, 1976). The pH was reduced with H_3PO_4 (75 %, Novadan ApS, Denmark). The pulses were performed in three replicates. Six pH sensors (CPS471D, Endress+Hauser AG, Switzerland) mounted on a lance measured the pH at different positions every second. The positions were equivalent with the monitoring points in the tracer pulse simulation with the exception of the top placed sensor, which failed to record the data (Fig. 1).

Mixing time calculation

Mixing times were calculated after normalizing the pH measurements according to Paul et al. (2003):

$$pH'_{i,exp}(t) = \frac{pH_{i.exp}(t) - pH_{i,exp}(t=0)}{pH_{i,exp}(t=\infty) - pH_{i,exp}(t=0)} \quad (2)$$

where $pH'_{i,exp}$ is the normalized pH output of the i-th sensor in the experiment, $pH_{i,exp}$ is the experimental pH value measurement, and $pH_{i,exp}(t=\infty)$ are the average pH measurements between 4.5 and 5 minutes after the pulse. The normalized response of all sensors was plotted with the logarithmic squared deviation with respect to the normalized upper bound 1 in order to determine the mixing time:

$$log\ D^2 = log\left[\frac{1}{n}\cdot\sum_{i=1}^{n}\left(pH'_{i,exp}(t)-1\right)^2\right] \quad (3)$$

where n is the number of sensors. $\log D^2 = -2.6$ when 95% homogeneity was achieved, $\log D^2 = -2$ and $\log D^2 = -1.65$ at 90% and 85 % homogeneity, respectively. The simulated tracer pulse concentrations were normalized by eq. (2), in which the pH values were replaced by the tracer concentrations.

***Streptococcus thermophilus* fermentation and analysis**

The batch fermentation of the homolactic *S. thermophilus* strain (provided by Chr. Hansen A/S, Hørsholm, Denmark) was carried out in the aforementioned 700 L stirred tank bioreactor at a stirring speed of 132 rpm, 40 °C, and N_2 headspace gassing. The pH was controlled by adding 24 % (w/v) ammonia solution (NH_4OH) through a pipe, which was located 0.1 m above the bottom in the center of the vessel, to maintain pH = 6.0. The pH-value was measured from a sensor, which was located 0.3 m above the bottom of the bioreactor close to the reactor wall. The initial pH was 6.8. The medium contained 70 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$. The pH was monitored every second at 4 of the 6 previously mentioned positions at heights of 0.10 m, 0.60 m, 1.25 m, and 1.60 m. Dry cell weight was determined from centrifuged, washed (with 0.9 % NaCl), and dried (at 70 °C for 24 h) cell broth. Sugars and organic acids were quantified from filtered (0.2 µm) samples in an HPLC system (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA) and a refractive index detector (ERC RefractoMax 520), with an Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA) using 5 mM H_2SO_4 at a flow rate of 0.6 mL min^{-1} at 50 °C according to suppliers instructions.

Biokinetic and pH simulation in the CFD model

An unstructured kinetic model of *S. thermophilus*, which described the lactose consumption, biomass growth, and lactic acid synthesis, was integrated into the CFD model (eq. (4-7)). Effects of the lag-time, substrate limitation and inhibition (Haldane, 1930), pH (Schepers et al., 2002), and lactate inhibition (Aghababaie et al., 2015) were considered in the growth function. A simplified version of the Luedeking-Piret equation (Luedeking and Piret, 1959) was applied to describe the lactic acid synthesis.

$$\left(1 + Y_{gal}\right) \cdot lactose \xrightarrow{q_X} biomass + Y_{gal} \cdot galactose \tag{4}$$

$$\left(1 + Y_{gal}\right) \cdot lactose \xrightarrow{q_P} lactic\ acid + Y_{gal} \cdot galactose \tag{5}$$

$$q_X = \mu_{max} \cdot \left(1 - e^{-t/t_{lag}}\right) \cdot \frac{C_S}{C_S + K_s + \frac{C_S^2}{K_I}} \cdot e^{-\left(\frac{\left(pH_{opt} - pH\right)^2}{\sigma^2}\right)} \cdot \frac{1}{1 + e^{K_{P,La}\left(C_P - \frac{K_{La}}{1 + e^{K_{P,pH1} \cdot \left(pH - K_{P,pH2}\right)}}\right)}} \cdot C_X \tag{6}$$

$$q_P = \alpha \cdot q_X \tag{7}$$

where q_X and q_P are the volumetric growth and lactic acid production rates, respectively. Lactose (C_S), biomass (C_X), and lactic acid (C_P) were listed as additional variables, and their rate equations were defined as expressions in the CFX expression language. Initial concentrations were $C_{S,t=0} = 70\ g\ L^{-1}$, $C_{X,t=0} = 0.025\ g\ L^{-1}$, and $C_{P,t=0} = 0\ g\ L^{-1}$. The kinetic parameters as listed in Table I were derived from a parameter estimation, which was based on 2 L lab-scale fermentations with the aforementioned medium at 300 rpm (two 6-blade Rushton turbines with a diameter = 53 mm) and 40 °C at different pH values (in the range of 5.5 – 7.0) and initial lactose concentrations (20 and 70 g L^{-1}) (Spann et al., 2018). It must be considered in the evaluation of the model that the supplemented yeast extract contains ca. 6 g L^{-1} carbon, which is not included in the model. However, this is only partially taken up by the cells and the dynamic model accounts for it by under predicting the galactose concentration. The biomass, lactic acid, and lactose concentrations, which are crucial in this study, are predicted accurately (Spann et al., 2018).

Table I. Kinetic parameters of the integrated *S. thermophilus* model.

Symbol	Description	Value
K_I	substrate inhibition parameter	164 g L^{-1}
K_S	substrate limitation parameter	0.79 g L^{-1}
K_{La}	lactate inhibition parameter	21.1 g L^{-1}
$K_{P,La}$	2. lactate inhibition parameter	0.2 L g^{-1}
$K_{P,pH1}$	LA inhibition pH parameter	20
$K_{P,pH2}$	2. LA inhibition pH parameter	7
pH_{opt}	optimal pH in the pH function	6.22
t_{lag}	lag-time coefficient	0.38 h
Y_{gal}	galactose yield	0.63 g g^{-1}
α	growth related production coefficient of lactic acid	5.59 g g^{-1}
μ_{max}	maximum specific growth rate	2.16 h^{-1}
σ	spread parameter in the pH function	1.09

An algebraic linear correlation for the pH calculation based on the lactic acid and ammonia concentrations was obtained based on experiments performed at 2 L scale (Supplementary Material):

$$pH = -0.44 \cdot \left(C_P - 5.29 \cdot C_{NH_3}\right) + 7.00 \quad (8)$$

The dynamic simulation with a time step of 1 s and an RMS residual target of $1 \cdot 10^{-5}$ was carried out using a steady state result as initialization state. Continuity, momentum, and energy equations were derived from a steady state solution, and thus assumed constant. They were therefore not solved in the dynamic simulation in order to reduce the computational time. The impeller speed was set to 200 rpm for the steady state velocity profile in the fermentation simulation in contrast to 132 rpm in the experimental fermentation. This modification was necessary in order to represent the mixing behaviour in the fermentation simulation (with a steady state velocity profile) as the predicted mixing times differed when applying a steady state or transient velocity profile (see the Results and the Discussion sections for details). The pH was controlled by adding ammonia at the same position as in the experiment. Control was conducted with a P-controller, which was using the step function:

$$NH_{3,add} = step(6 - pH) \cdot (6 - pH) \cdot 11900\ g\ h^{-1} \quad (9)$$

where the pH is calculated at the monitoring point 35 cm above the bottom of the bioreactor. The kinetic model was also implemented in MATLAB (The MathWorks, Natick, MA) and solved with the ode 15s solver. There, the fermentation broth was modelled as a single compartment with a homogeneous distribution of the pH and all state variables, i.e. no gradients were considered.

Results

A one-phase CFD model of a 700 L bioreactor for a *S. thermophilus* fermentation was set up and tracer pulse experiments with NaOH and multi-position pH monitoring validated the fluid dynamic model predictions. A kinetic model of *S. thermophilus* was integrated into the validated CFD model in order to predict pH gradients during the fermentation.

The velocity profile of the bioreactor

A steady state solution of the CFD model was initially obtained, which predicted the macroscopic flow profile of the bioreactor. It clearly revealed six recirculation loops, which were generated by the Rushton turbines (Fig. 2). A turbulent flow regime was assumed, because the Reynolds number was $2.2 \cdot 10^5$ at 242 rpm. The fluid velocity was highest behind the turbine blades, which turned with 2.8 m s^{-1} tip speed at 242 rpm. Low velocities were observed close to the bioreactor wall and especially around the baffles. The steady state solution converged with respect to the RMS values of the velocity components, while the velocities were unstable at the monitoring points. Further analysis revealed that the velocity profile of the bioreactor had a transient (time-dependent) nature (Supplementary Movie 1). The four recirculation loops *between* the impellers were changing in size and moving up- and downwards. The tracer pulse simulation was therefore performed with a transient velocity field.

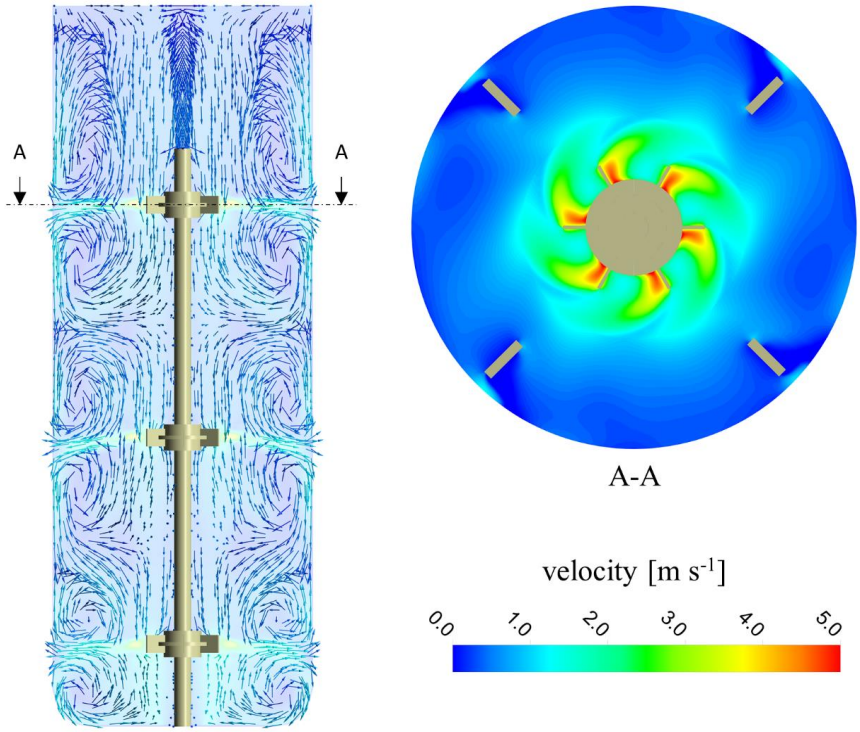


Figure 2. Steady state solution of the 700 L stirred tank bioreactor with 300 rpm. Left: velocity streamlines with velocity in stationary frame. Right: contour plot with the circumferential velocity in stationary frame.

Tracer pulse simulation and experiments

Fast radial and slower axial mixing were predicted in the tracer pulse simulations (Fig. 3 and Supplementary Movie 2). It took several seconds until the tracer passed to a subsequent recirculation loop after the simulated injection from the top. The monitoring points in the CFD model and pH sensors in the experiment were positioned in each anticipated recirculation loop in order to be able to follow the dynamic distribution of the tracer.

The dynamic response at all sensor locations

The dynamic responses of the monitoring points during tracer pulse simulations were captured with the intention to understand the fluid flow dynamics when e.g. an acid or a base solution is added to regulate the pH value in a cultivation. The two monitoring locations 1 and 2 at the upper part of the liquid phase showed an overshooting response before they reached a stable value, whereas the other monitoring points 3 – 6, which were located in a greater distance to the injection point, responded with sigmoid curves (Fig. 4). In order to validate the tracer pulse simulations, these results were compared with the experimental measurements. Both the shapes and order of magnitude of the dynamic trends obtained from the predictions agreed with the measurements at the different positions. Nevertheless, oscillations of the pH signal and the initial high overshoot of sensor 2 predicted by the simulation were not captured by the measurements.

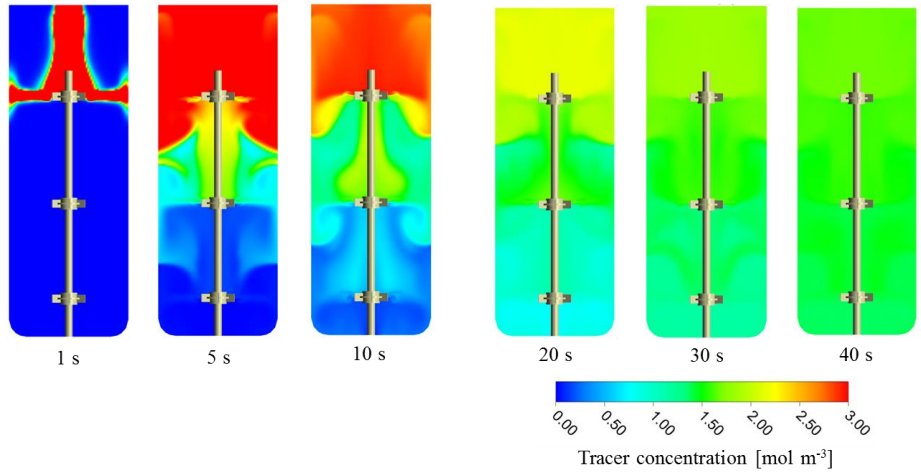


Figure 3. Concentration fields of the tracer during the transient simulation of the pulse in the 700 L stirred tank bioreactor at 242 rpm. The tracer was injected at 0 s at the top of the liquid phase and snapshots are taken at different time points.

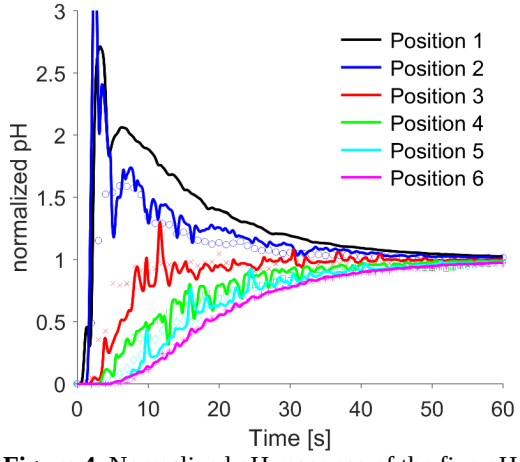


Figure 4. Normalized pH response of the five pH sensors (Position 2-6, as shown in Fig. 1) in the pulse experiment and six monitoring points in the transient simulation performed at 242 rpm. Experimental values (symbols) and simulated values (solid lines). The pH showed an overshoot close to the injection point at the top of the bioreactor before it leveled out. The pH increased gradually at the lower positioned sensors and monitoring points.

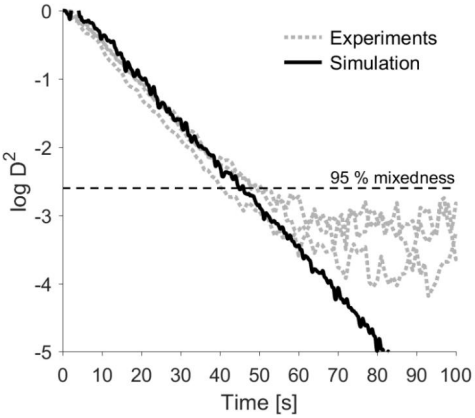


Figure 5. Logarithmic squared deviation of the pH values in the tracer pulse experiments and simulation considering all monitor points. Three tracer pulse experiments (dotted lines) and the CFD simulation (solid line) at 242 rpm. 95 % homogeneity was reached at log D^2 = -2.6.

Determination of the mixing time

To assess the progress of reaching homogeneity, the logarithmic squared deviation of all sensors was evaluated. All experimental curves followed the predicted trend until 95 % homogeneity was achieved (Fig. 5). The variance of the replicates increased for homogeneities higher than 95 %. The predicted and measured mixing times at the levels of 85 %, 90 %, and 95 % homogeneity matched very well (Table II). 95 % homogeneity was reached after about 46 seconds.

The dynamic response at all locations and the mixing time prediction gave considerable evidence that the fluid flow in the bioreactor was well described by the applied CFD model.

Table II. Experimental and CFD predicted mixing times for different levels of homogeneity at 242 rpm (P/V = 0.79 kW m^{-3}).

Level of homo-geneity	Mixing time for the tracer pulses [s]	
	Experiments	CFD simulation
85 %	26 30 30	30
90 %	32 36 35	36
95 %	42 50 51	46

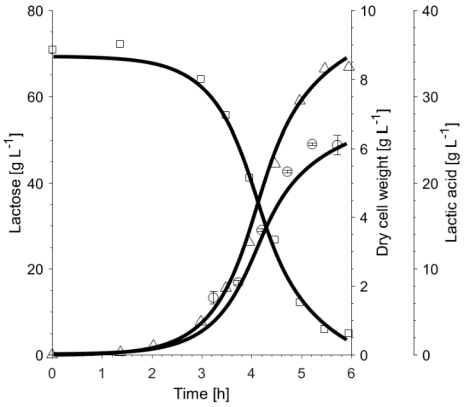


Figure 6. Dry cell weight, lactose, and lactic acid concentrations as measured and predicted in the *S. thermophilus* batch fermentation. The fermentation was carried out in the 700 L bioreactor at 132 rpm, 40 °C, and the pH controlled at pH = 6. Dry cell weight (circles) with standard deviation, lactose (squares), lactic acid (triangles), and the CFD simulation result (solid line).

Simulated and measured pH gradients in the *S. thermophilus* fermentation

As a next step, the CFD and biokinetic models were combined in a dynamic simulation in order to predict the pH gradients during the fermentation. Therefore, an unstructured non-segregated kinetic model of *S. thermophilus* and a pH correlation were integrated into the validated CFD model. A dynamic simulation was performed with the purpose of predicting the pH gradients during the batch fermentation. A steady state velocity profile was applied, which decreased the computational demand in contrast to solving the fluidic profile for the entire fermentation time. The mixing time at the fermentation conditions of 132 rpm was 85 s according to a tracer pulse simulation with a transient velocity profile. The impeller speed had to be set to 200 rpm in the fermentation simulation in order to represent the same mixing time with a steady state velocity profile (Supplementary Fig. S1). Expected biomass growth, substrate consumption, and lactic acid production profiles of a Monod type kinetic model were observed and in accordance with the measurements (Fig. 6). A final biomass concentration of 6 g L^{-1} was reached after 5 h when 34 g L^{-1} lactic acid seemed to inhibit growth completely. The observed biomass yield was similar to 2 L lab-scale experiments, where perfect mixing was assumed (Supplementary Fig. S2). As the applied time step was crucial in order to solve the differential equations in the CFD model, a time step of 1 s was chosen. An increased time step led to larger deviations of the kinetic profiles (data not shown). The obtained kinetic results from the CFD simulation were very similar to the single-compartment simulation performed in MATLAB, in which a homogeneous distribution of the pH and all state variables was assumed (Supplementary Fig. S3). The pH predictions were in close agreement with the measurements in all locations (Fig. 7). In the beginning of the fermentation, the pH dropped from 6.8 to the controlling pH value 6, when the pH controller started in both the simulation and experiment. A minimum pH of 5.9 was attained in the top zone of the bioreactor during the exponential growth phase. In the bottom zone, where ammonium hydroxide was added, pH values of up to 6.3 were measured and simulated at the sensor positions. Close to the base injection, pH values larger than 7 were predicted (Fig. 8). As the applied pH correlation is only valid up to a pH of 7, a more accurate pH prediction was not possible in this case.

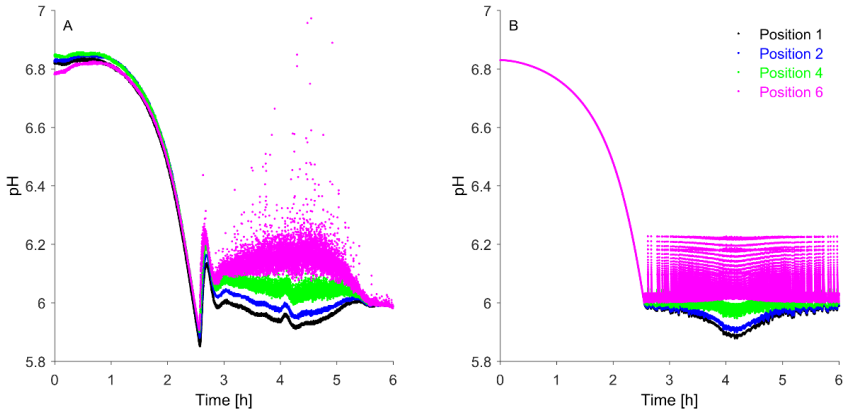


Figure 7. pH values measured and predicted every second at different positions during the *S. thermophilus* fermentation. Fermentation (A) and CFD simulation (B). In the beginning of the fermentation, the pH dropped from 6.8 to the controlling pH value 6, when the pH controller started. The pH was controlled at pH = 6 using the measurement of sensor 5 by adding NH_4OH at the bottom of the bioreactor. pH sensors and monitoring points were placed at position 1, 2, 4, and 6 in the bioreactor as shown in Fig. 1. The pH dropped down to pH = 5.9 in the top zone of the bioreactor, whereas a maximum of around pH = 6.3 was measured and predicted at position 6 in the bottom zone of the bioreactor.

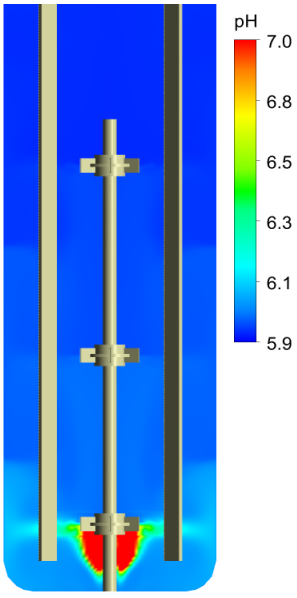


Figure 8. Simulated pH gradients during the *S. thermophilus* fermentation in the 700 L bioreactor at 4 h 40 min of the cultivation. The pH was higher than 7 close to the alkali inlet at the bottom of the bioreactor and around 5.9 in the top zone of the bioreactor. As the applied pH correlation is only valid up to a pH of 7, a more accurate pH prediction was not possible.

Discussion

In order to simulate the pH gradients of a *S. thermophilus* fermentation in a 700 L bioreactor, a one-phase CFD model was first validated and then coupled with a biokinetic model and a pH correlation. Multi-position pH monitoring in tracer pulse experiments validated the fluid dynamic model predictions of the one-phase CFD model. The CFD model predicted the mixing time of around 46 s to reach 95 % homogeneity at an impeller speed of 242 rpm and forecasted the dynamic response of all sensors in the tracer pulse experiments. The dynamic simulation of the non-aerated *S. thermophilus* batch fermentation predicted both the biokinetic profiles and the pH gradients matching the experimental observations. Rather large pH gradients between pH values of 5.9 and higher than 7 were predicted in the bioreactor while the fermentation was controlled at pH 6.

The simulated flow profiles showed six recirculation loops generated by the three Rushton turbines consistent with literature data (Vrabel et al., 2000). pH sensors and monitoring points were placed so that conditions in all six recirculation loops were monitored in tracer pulse experiments and simulated accordingly. The dynamic pH response of the pH sensors was well represented by the simulated data. It is important to stress that the CFD model relied among other criteria on physical and chemical properties, empirical equations, and the mesh structure. Importantly, no parameter estimation/model calibration of the CFD model was performed in order to fit experimental data. However, the predicted oscillating behavior of the pH and the initial overshoot of sensor 2 was not measured. This can likely be attributed to the response time of the applied ISFET pH sensors, which is in a range of 4 - 8 s to reach ±0.02 of the final pH value in the relevant pH range. This response time was determined in own measurements, and is in accordance with vendor specifications. Furthermore, there was a discrepancy between the predicted and measured homogeneity when 95 % homogeneity was reached 60 s after the pulse, which could be caused by the fluctuating sensor output (±0.01), while the model asymptotically approaches 100 % homogeneity.

It was shown that the recirculation loops were dynamically changing, and hence a transient velocity profile was required. Dynamic velocity changes that might have caused the dynamic behaviour of the recirculation loops have been already observed for Rushton turbines (Hartmann et al., 2004; Nikiforaki et al., 2003). However, the velocities have not been experimentally validated for the studied system, yet.

Both observed and simulated mixing times were consistent with results from Delvigne et al. (2006). They reported similar mixing times between 20 and 53 s to reach 85 % homogeneity in stirred tank bioreactors with a working volumes of 350, 1200, and 1800 L with a comparable power input to the present study. However, as no power input measurements were available for our 700 L bioreactor, the theoretical power input could not be validated in this study. With regard to the definition of mixing time in CFD simulations, Larsson (2015) concluded that there is no consistent definition so far. Instead, there exist several possibilities to determine the mixing time from observing the CFD system at one or several points, up to detecting the concentration on flat surface planes covering a larger area of the CFD system. In contrast to previous studies, which only used one position to calculate the mixing time, six points, which were distributed over the whole liquid phase, were used in this study in order to improve reproducibility and accuracy of the results. Overall, it should be noted that both the experimental setup as well as the way of treating and interpreting the data still lead to uncertainties. For example, the location of the top sensors and their monitoring points affects the accuracy of the measurements and predictions of the overshoot after the tracer pulse. A sensitivity analysis of the sensor locations in the simulation could support the assessment of

the accuracy of the model. Up to now, this study has proven that the CFD model achieved a good prediction of the fluid dynamics in the bioreactor.

Discussion of the results from the combined CFD and biokinetic model

Since heterogeneities at large scale affect the productivity of many chemical and biochemical processes, a tool to couple fluid dynamics and reaction kinetics is highly demanded. Dynamic simulations of biokinetic models integrated in the fluidic profile simulated by a CFD model can pave the way for enhanced understanding of microbial behavior in large-scale bioreactors. Consequently, it is a basic requirement that the CFD simulation provides accurate results.

To achieve an affordable computational time for the biokinetic CFD simulation, a steady state velocity profile was required. The necessary manipulation of the stirrer speed (to 200 rpm) in the steady state simulation was necessary because of the general transient fluid dynamics in the bioreactor as discussed above. It could also be considered to apply other turbulent models in future. However, a tuning of the CFD model to fit the experimental data should be generally avoided, and the computational development might allow in future using the transient velocity profile for the biokinetic CFD simulation.

Due to the higher computational demand, while solving the differential equations in all nodes of the CFD model mesh (about 1.6 million nodes in this study), discretization errors are likely when selected time steps are inappropriate. The same issue occurs if RMS targets are too high. Applying a time step of 1 s resulted in a similar biological growth as observed in the experiment, while larger time steps led to larger deviations between measurements and predictions. This is most likely due to the accumulation of numerical errors. However, a smaller time step might have reduced numerical errors further, but will also increase computational burden. The similarity between the single-compartment simulation – where completely mixed conditions were assumed – and the CFD simulation results might be caused by three reasons: (i) either the pH gradients had a very small effect on the culture performance in the present study, or (ii) the biokinetic model was not sensitive to pH changes, or (iii) the small differences arose from the aforementioned numerical errors in the CFD simulation due to the coarse time step.

pH gradients between 5.9 and 6.3 were predicted and observed between the top and the bottom zone of the bioreactor, respectively. A pH higher than 7 was predicted for the vicinity of the base solution inlet. Even though the pH measurements and predictions matched, the uncertainties in the applied pH correlation need to be considered. The fast production of lactic acid led to a small decrease of the pH at the top of the bioreactor, whereas the addition of ammonium hydroxide caused a pulse-wise increase of the pH at the bottom of the bioreactor. Langheinrich and Nienow (1999) reported pH gradients of 0.8 units due to alkali addition in an 8 m^3 reactor for mammalian cell cultures. pH gradients have a noticeable effect on the productivity. Aghababaie et al. (2015) reported that the growth of *S. thermophilus* was reduced by 20 % when cultivated 0.3 pH units away from the optimal pH conditions. However, cells are not constantly exposed to unfavorable environmental conditions while moving through a large-scale bioreactor. Cortés et al. (2016) and Amanullah et al. (2001) showed in two-compartment scale-down studies of *E. coli* and *B. subtilis*, in which they mimicked oscillating pH conditions up to a delta pH of 0.9, that growth was not statistically significantly affected. However, the organic acid metabolism changed, and *E. coli* responded on the transcriptional level to the alkaline stress. The extracellular pH affects the intracellular pH of lactic acid bacteria (Cachon et al., 1998; Hansen et al., 2016) and by this the enzymatic activity. *Lactobacillus* sp. maintain their intracellular pH with the energy consuming Na^+ (K^+)/H^+ antiporters (Sawatari and Yokota, 2007). The additional energy requirements could lead to altered culture performance in large-scale fermentations. The remaining open question is how

fast the cells are affected by pH changes and how fast they adapt to them. In the immediate vicinity of the inlet for base addition, the cells are exposed to unfavorable pH values that might lead to viability loss (Hansen et al., 2016). Cells that are moving through the bioreactor and have suffered in an unfavorable environment before will not function in an optimal manner immediately, when they enter a more favorable zone, as they need to adapt to the new conditions again (Löffler et al., 2016; Nieß et al., 2017). Further studies like Vanrolleghem et al. (2004), who studied and modelled the dynamic response to substrate pulses in wastewater treatment plants, are required to understand the adaptation processes of microorganisms under oscillating conditions better. This knowledge about metabolic phenomena, e.g. the dynamic response of growth to changing substrate availability, could then expand the biokinetic models coupled with CFD models (Delvigne and Noorman, 2017).

Coupling biokinetic and fluid dynamic modelling will open the way for the understanding and optimization of large-scale processes. To predict gradients at large scale is of utmost interest, because measurements during fermentations at large scale are either very difficult or even impossible to perform due to the size of the bioreactors, the costs of a single fermentation run, and the GMP regulations at production sites. Scale-down systems (Oosterhuis, 1984) could be designed based on the CFD predictions, and mimic the gradients in lab-scale experiments (Lara et al., 2006; Neubauer and Junne, 2016). They allow to study the response mechanisms upon external oscillating conditions, and can be integrated in the scale-up process (Neubauer et al., 2013). This will reduce the risk of failure when scaling up processes.

Conclusion

The present study was designed to predict pH gradients in a 700 L lactic acid bacteria fermentation by applying a dynamic CFD simulation. It gave evidence that pH heterogeneities existed in the studied 700 L bioreactor. More importantly, it proved that pH gradients could be quantitatively predicted with the CFD simulation. pH gradients between 5.9 at the top and above 7 close to the alkali inlet at the bottom of the bioreactor were predicted. The high pH in the alkali inlet zone could cause cell damage and an undesired production loss in large-scale bioreactors. Therefore, the results could support fine-tuning of the stirring rate when reaching the maximum growth rate in order to distribute the base faster. In summary, these findings suggest that coupling a biokinetic model and a fluid dynamic model is a very useful tool to predict gradients in bioreactors. However, to predict the effect on microorganisms growing under oscillating conditions was beyond the scope of this study. The validation of the applied CFD model with multi-position pH monitoring during mixing experiments is a promising outcome of this study, which should be performed in further CFD studies of bioprocesses as well. Future work should include multiple environmental gradients in the dynamic CFD simulations. Besides pH, also substrate, oxygen, carbon dioxide, and temperature gradients are of high interest for batch, fed-batch, and continuous cultivations since most of them are regarded to contribute to productivity loss at large scale.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056. The authors would like to thank Jifeng Yang, Pinxiang Han, and Peter Nielsen for the support during this study. We are grateful to Chr. Hansen A/S for the cooperation.

References

Aghababaie M, Khanahmadi M, Beheshti M. 2015. Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures. *Food Bioprod. Process.* **94**:657–667.

Amanullah A, McFarlane CM, Emery AN, Nienow AW. 2001. Scale-down model to simulate spatial pH variations in large-scale bioreactors. *Biotechnol. Bioeng.* **73**:390–399.

Bannari R, Bannari A, Vermette P, Proulx P. 2012. A model for cellulase production from Trichoderma reesei in an airlift reactor. *Biotechnol. Bioeng.* **109**:2025–2038.

Bylund F, Collet E, Enfors S-O, Larsson G. 1998. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Eng.* **18**:171.

Cachon R, Antérieux P, Diviès C. 1998. The comparative behavior of Lactococcus lactis in free and immobilized culture processes. *J. Biotechnol.* **63**:211–218.

Cents AHG, Brilman DWF, Versteeg GF. 2005. CO2 absorption in carbonate/bicarbonate solutions: The Danckwerts-criterion revisited. *Chem. Eng. Sci.* **60**:5830–5835.

Cortés JT, Flores N, Bolívar F, Lara AR, Ramírez OT. 2016. Physiological effects of pH gradients on Escherichia coli during plasmid DNA production. *Biotechnol. Bioeng.* **113**:598–611.

Delvigne F, Destain J, Thonart P. 2006. A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. *Biochem. Eng. J.* **28**:256–268.

Delvigne F, Noorman H. 2017. Scale-up/Scale-down of microbial bioprocesses: a modern light on an old issue. *Microb. Biotechnol.* **10**:685–687.

Doran PM. 1995. Bioprocess Engineering Principles. London, UK: Academic Press.

Einsele A. 1976. Charakterisierung von Bioreaktoren durch Mischzeiten. *Chem. Rundschau* **25**:53–55.

Elqotbi M, Vlaev SD, Montastruc L, Nikov I. 2013. CFD modelling of two-phase stirred bioreaction systems by segregated solution of the Euler-Euler model. *Comput. & Chem. Eng.* **48**:113–120.

Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägårdh C, Fuchs L, Revstedt J, Friberg PC, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, Van der Lans R, Luyben K, Vrabel P, Manelius Å. 2001. Physiological responses to mixing in large scale bioreactors. *J. Biotechnol.* **85**:175–185.

George S, Larsson G, Olsson K, Enfors S-O. 1998. Comparison of the Baker's yeast process performance in laboratory and production scale. *Bioprocess Eng.* **18**:135.

Haldane JBS. 1930. Enzymes. *Enzymes*.

Hansen G, Johansen CL, Marten G, Wilmes J, Jespersen L, Arneborg N. 2016. Influence of extracellular pH on growth, viability, cell size, acidification activity, and intracellular pH of Lactococcus lactis in batch fermentations. *Appl. Microbiol. Biotechnol.* **100**:5965–5976.

Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ, Mudde RF, Noorman HJ. 2016. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Eng. Life Sci.* **16**:652–663.

Hartmann H, Derksen JJ, Van Den Akker HEA. 2004. Macroinstability uncovered in a Rushton turbine stirred tank by means of LES. *AIChE J.* **50**:2383–2393.

Langheinrich C, Nienow AW. 1999. Control of pH in large-scale, free suspension animal cell bioreactors: Alkali addition and pH excursions. *Biotechnol. Bioeng.* **66**:171–179.

Lapin A, Müller D, Reuss M. 2004. Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler−Lagrange Methods: Traveling along the Lifelines of Single Cells †. *Ind. Eng. Chem. Res.* **43**:4647–4656.

Lapin A, Schmid J, Reuss M. 2006. Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor-A structured-segregated approach. *Chem. Eng. Sci.* **61**:4783–4797.

Lara AR, Galindo E, Ramírez OT, Palomares L a. 2006. Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol. Biotechnol.* **34**:355–381.

Larsson G, Törnkvist M, Ståhl Wernersson E, Trägårdh C, Noorman H, Enfors SO. 1996. Substrate gradients in bioreactors: Origin and consequences. *Bioprocess Eng.* **14**:281–289.

Larsson HK. 2015. Modelling of Mass Transfer Phenomena in Chemical and Biochemical Reactor Systems using Computational Fluid Dynamics. *PhD thesis, Tech. Univeristy Denmark*.

Löffler M, Simen JD, Jäger G, Schäferhoff K, Freund A, Takors R. 2016. Engineering E. coli for large-scale production – Strategies considering ATP expenses and transcriptional responses. *Metab. Eng.* **38**:73–85.

Luedeking R, Piret EL. 1959. Kinetic study of the lactic acid fermentation. Batch process at controlled pH. *Biotechnol. Bioeng.* **67**:636–644.

Morchain J, Gabelle J-C, Cockx A. 2013. Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. *AIChE J.* **59**:369–379.

Nagata S. 1975. Mixing: Principles and Application. New York: Wiley.

Neubauer P, Cruz N, Glauche F, Junne S, Knepper A, Raven M. 2013. Consistent development of bioprocesses from microliter cultures to the industrial scale. *Eng. Life Sci.* **13**:224–238.

Neubauer P, Junne S. 2016. Scale-Up and Scale-Down Methodologies for Bioreactors. In: Mandenius, CF, editor. *Bioreact. Des. Oper. Nov. Appl.* Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 323–354.

Nieß A, Löffler M, Simen JD, Takors R. 2017. Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of E. coli cultures in large-scale bioreactors: Experimental evidence and mathematical model. *Front. Microbiol.* **8**:1–9.

Nikiforaki L, Montante G, Lee KC, Yianneskis M. 2003. On the origin, frequency and magnitude of macro-instabilities of the flows in stirred vessels. *Chem. Eng. Sci.* **58**:2937–2949.

Onyeaka H, Nienow AW, Hewitt CJ. 2003. Further Studies Related to the Scale-up of High Cell Density Escherichia coli Fed-Batch Fermentations: The Additional Effect of a Changing Microenvironment When Using Aqueous Ammonia to Control pH. *Biotechnol. Bioeng.* **84**:474–484.

Oosterhuis N. 1984. Scale-up of bioreactors: a scale-down approach. *PhD thesis, Delft Univ. Technol.*

Paul EL, Atiemo-Obeng VA, Kresta SM eds. 2003. Handbook of Industrial Mixing. *Handb. Ind. Mix. Sci. Pract.* Hoboken, NJ, USA: John Wiley & Sons, Inc.

Ruston JH, Costich EW, Everett HJ. 1950. Power Characteristics of Mixing Impellers Part II. *Chem. Eng. Prog.* **46**:467–476.

Sawatari Y, Yokota A. 2007. Diversity and Mechanisms of Alkali Tolerance in Lactobacilli. *Appl. Environ. Microbiol.* **73**:3909–3915.

Schepers AW, Thibault J, Lacroix C. 2002. Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. *Enzyme Microb. Technol.* **30**:187–194.

Schmalzriedt S, Jenne M, Mauch K, Reuss M. 2003. Integration of Physiology and Fluid Dynamics. In: von Stockar, U, van der Wielen, LAM, Bruggink, A, Cabral, JMS, Enfors, S-O, Fernandes, P, Jenne, M, Mauch, K, Prazeres, DMF, Reuss, M, Schmalzriedt, S, Stark, D, von Stockar, U, Straathof, AJJ, editors. *Process Integr. Biochem. Eng.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–68. http://dx.doi.org/10.1007/3-540-36782-9_2.

Spann R, Roca C, Kold D, Eliasson Lantz A, Gernaey K V., Sin G. 2018. A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. *Biochem. Eng. J.* **135**:49–60.

Vanrolleghem PA, Sin G, Gernaey K V. 2004. Transient response of aerobic and anoxic activated sludge activities to sudden substrate concentration changes. *Biotechnol. Bioeng.* **86**:277–290.

Vrábel P, Van der Lans RGJM, Van der Schot FN, Luyben KCAM, Xu B, Enfors SO. 2001. CMA: Integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chem. Eng. J.* **84**:463–474.

Vrabel P, Van Der Lans RGJM, Luyben KCAM, Boon L, Nienow AW. 2000. Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: Modelling and measurements. *Chem. Eng. Sci.* **55**:5881–5896.

Wang G, Tang W, Xia J, Chu J, Noorman H, van Gulik WM. 2015. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Eng. Life Sci.* **15**:20–29.

Xu B, Jahic M, Blomsten G, Enfors S-O. 1999. Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with Escherichia coli. *Appl. Microbiol. Biotechnol.* **51**:564–571.

Yang Y, Xia J, Li J, Chu J, Li L, Wang Y, Zhuang Y, Zhang S. 2012. A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *J. Biotechnol.* **161**:250–256.

Zou X, Xia J, Chu J, Zhuang Y, Zhang S. 2012. Real-time fluid dynamics investigation and physiological response for erythromycin fermentation scale-up from 50 L to 132 m3 fermenter. *Bioprocess Biosyst. Eng.* **35**:789–800.

Supplementary material

Contents

1. Supplementary Figures

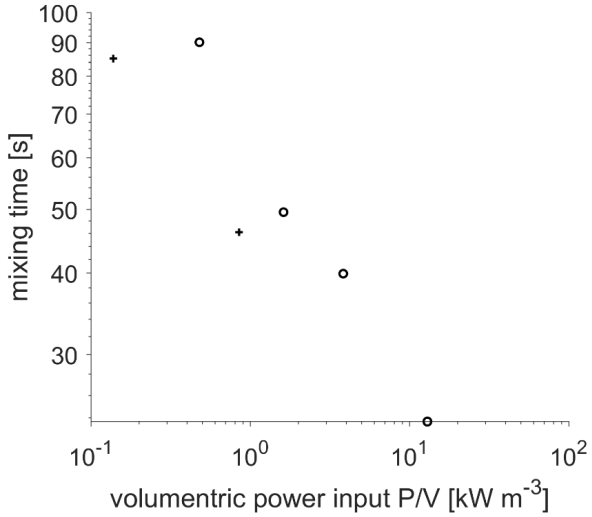


Figure S1. Mixing time prediction to reach 95 % homogeneity by the CFD simulation with different volumetric power inputs. Steady-state velocity profile (circles) and transient velocity profile (plus).

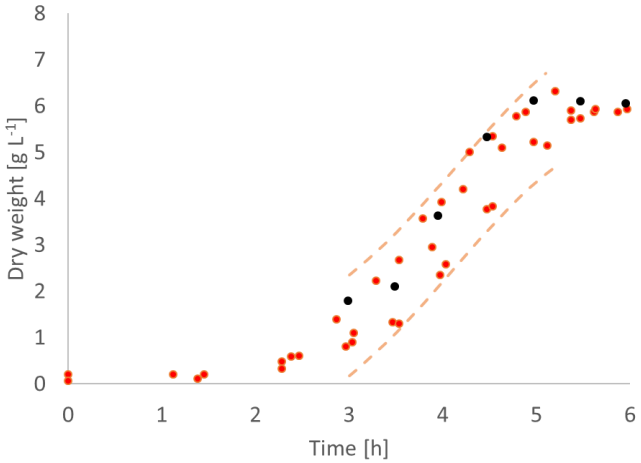


Figure S2. Comparison of the biomass growth in the 700 L bioreactor (black dots) and the 2 L bioreactor (red dots) with 95 % confidence interval of the three experiments in 2 L.

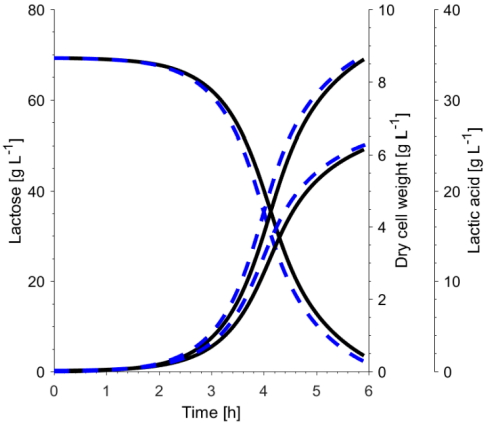


Figure S3. Comparison of the kinetic model results from the CFD and MATLAB simulation. CFD (solid black line) and MATLAB (dashed blue line) simulation. The bioreactor was assumed to be one element with homogeneous distribution of the pH and all state variables in the MATLAB simulation in contrast to the CFD simulation.

2. The k-epsilon model in ANSYS CFX

The k-epsilon model is applied in this study as it is implemented in ANSYS CFX 17.1 (see chapter 2.2.2.1. The k-epsilon Model in ANSYS CFX in the CFX-Solver Theory Guide). An overview of the applied settings in given in Table SI.

The continuity equation is:

$$\frac{\delta \rho}{\delta t} + \frac{\delta}{\delta x_j}\left(\rho U_j\right) = 0 \tag{1}$$

The momentum equation is:

$$\frac{\delta \rho U_i}{\delta t} + \frac{\delta}{\delta x_j}\left(\rho U_i U_j\right) = -\frac{\delta p'}{\delta x_i} + \frac{\delta}{\delta x_j}\left[\mu_{eff}\left(\frac{\delta U_i}{\delta x_j} + \frac{\delta U_j}{\delta x_i}\right)\right] + S_M \tag{2}$$

where S_M is the sum of the body forces (e.g. ρg if gravity is the only momentum source), μ_{eff} is the effective viscosity accounting for turbulence, and p' is the modified pressure.

$$p' = p + \frac{2}{3}\rho k + \frac{2}{3}\mu_{eff}\frac{\delta U_k}{\delta x_k} \tag{3}$$

$$\mu_{eff} = \mu + \mu_t \tag{4}$$

where μ_t is the turbulence viscosity.

$$\mu_t = 0.09\rho\frac{k^2}{\varepsilon} \tag{5}$$

k is the turbulence kinetic energy and ε the turbulence eddy dissipation. k and ε derive from the differential transport equations for the turbulence kinetic energy and turbulence dissipation rate without buoyancy turbulence:

$$\frac{\delta \rho k}{\delta t} + \frac{\delta}{\delta x_j}(\rho U_i k) = \frac{\delta}{\delta x_j}\left[(\mu + \mu_t)\frac{\delta k}{\delta x_j}\right] + P_k - \rho\varepsilon \tag{6}$$

$$\frac{\delta \rho \varepsilon}{\delta t} + \frac{\delta}{\delta x_j}(\rho U_i \varepsilon) = \frac{\delta}{\delta x_j}\left[\left(\mu + \frac{\mu_t}{1.3}\right)\frac{\delta \varepsilon}{\delta x_j}\right] + \frac{\varepsilon}{k}(1.44P_k - 1.92\rho\varepsilon) \tag{7}$$

P_k is the turbulence production due to viscous forces:

$$P_k = \mu_t\left(\frac{\delta U_i}{\delta x_j} + \frac{\delta U_j}{\delta x_i}\right)\frac{\delta U_i}{\delta x_j} \tag{8}$$

List of symbols for the k-epsilon model

Symbol	Description	Dimension
g	gravity vector	m s^{-2}
k	turbulence kinetic energy per unit mass	m^2 s^{-2}
P_k	shear production of turbulence	g m^{-1} s^{-3}
p	static pressure	g m^{-1} s^{-2}
p’	modified pressure	g m^{-1} s^{-2}
S_M	momentum force	g m^{-2} s^{-2}
U	vector of velocity in the direction of x,y,z	m s^{-1}
ε	turbulence dissipation rate	m^{-2} s^{-3}
μ	dynamic viscosity	g m^{-1} s^{-1}
μ_t	turbulent viscosity	g m^{-1} s^{-1}
μ_{eff}	effective viscosity	g m^{-1} s^{-1}
ρ	density	g m^{-3}

Table SI. Overview of the settings for the CFD simulation.

Zones and settings	Value
Analysis type	Transient
	Timestep: Tracer pulse: 500 · 0.01 s following 0.1 s; Fermentation: 1 s
Reactor Domain	Material: Water
	Domain motion: Stationary
	Buoyancy model: Non Bouyant
	Turbulence option: k-Epsilon
	Heat transfer: None
Impeller Domains	Material: Water
	Domain motion: Stationary
	Buoyancy model: Non Bouyant
	Angular velocity: • Tracer pulse: 242 rpm (transient velocity profile) • Fermentation: 200 rpm (steady state velocity profile)
	Rotation axis: Global Y
	Turbulence option: k-Epsilon
	Heat transfer: None
Boundary conditions	Impeller top: Opening
	Reactor walls: No slip wall
	Reactor top: Free slip wall
	Reactor baffles: No slip wall
	Reactor shaft: No slip wall, rotating wall around the Y-axis
	Impeller surface: No slip wall
Interfaces	Rotational periodicity around the Y-axis between Impeller symmetry side 1 and 2
	Rotational periodicity around the Y-axis between Reactor symmetry side 1 and 2
	Top of the Impeller to the Reactor, Frame Change: Frozen Rotor
	Sides of the Impeller to the Reactor, Frame Change: Frozen Rotor
	Bottom of the Impeller to the Reactor, Frame Change: Frozen Rotor

3. Mesh study

A mesh size study was performed to find a mesh that provides accurate results. To achieve this the relative size of the mesh elements – and hence the number of nodes – was changed, and the simulated velocity profiles were assessed by comparing coarse and fine meshes. A steady state velocity profile was analyzed with the settings given in Table SI with 200 rpm.

First the rotating impeller domain was analyzed. Four different mesh sizes with 102k, 213k, 307k, and 487k nodes were evaluated. Besides a visual (subjective) interpretation (data not shown), the velocity along the line indicated in Fig. S4 was assessed. There were larger differences observed between the 102k mesh and the other simulations (Fig. S5). Smaller differences were seen between the 213k mesh and the finer meshes especially further away from the impeller blade. The differences between the 307k and the very fine 487k mesh were deemed to be small, and the 307k mesh was therefore selected for the impeller domain.

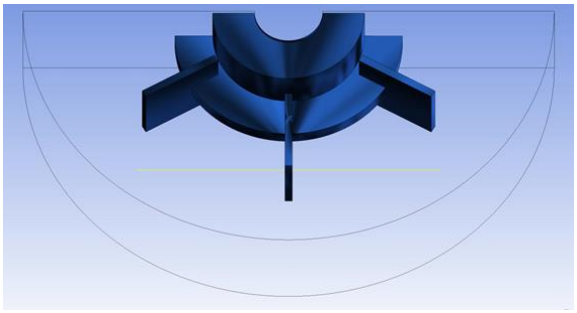


Figure S4. Line used to assess the optimum mesh size of the impeller domain.

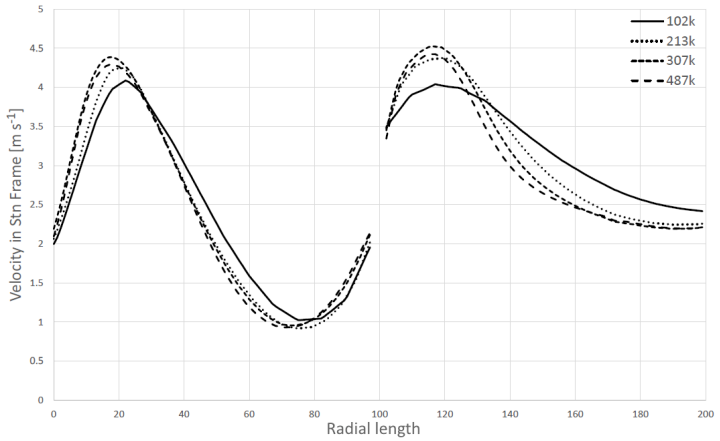


Figure S5. Velocity in the impeller domain for four different mesh sizes across the line in Fig. S4.

The mesh size of the bioreactor domain was subsequently analyzed with the above defined impeller domain (307k). Three mesh sizes were evaluated for the bioreactor domain, namely 475k, 712k, and 1008k nodes. The velocities were visual and systematically assessed along the horizontal and vertical line that are shown in Fig. S6.

While the velocity profiles were similar along the height of the bioreactor for all three cases, there were differences observed along the horizontal line (Fig. S7). The prediction with the coarse 475k mesh differed from the two finer meshes. The two finer meshes (712k and 1008k) however predicted similar velocity profiles. The 712k mesh was therefore considered further in this study as it was deemed to predict the velocity profile sufficiently, and had a lower computational demand than the 1008k mesh.

Nevertheless, the mesh size study must be interpreted with caution as firstly, the predicted velocities were not experimentally validated, and secondly, the steady state simulation that was

used in this mesh size study might not be suitable to describe this system entirely as the transient changing velocity was observed and discussed for this system in the main manuscript.

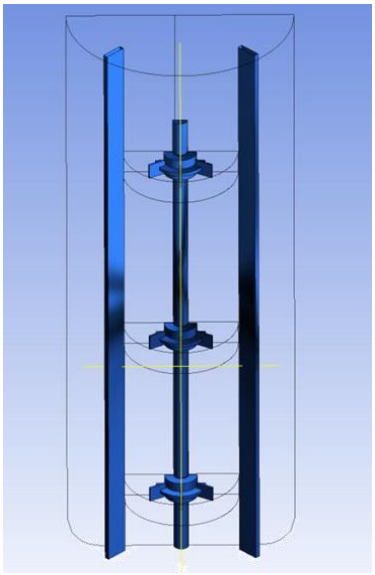


Figure S6. Horizontal and vertical lines used to assess the optimum mesh size of the bioreactor domain.

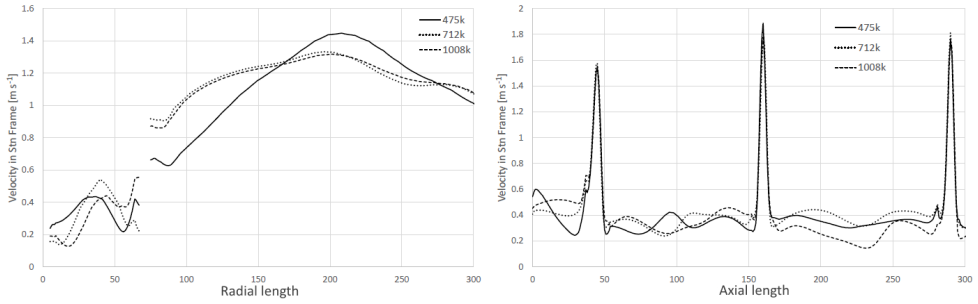
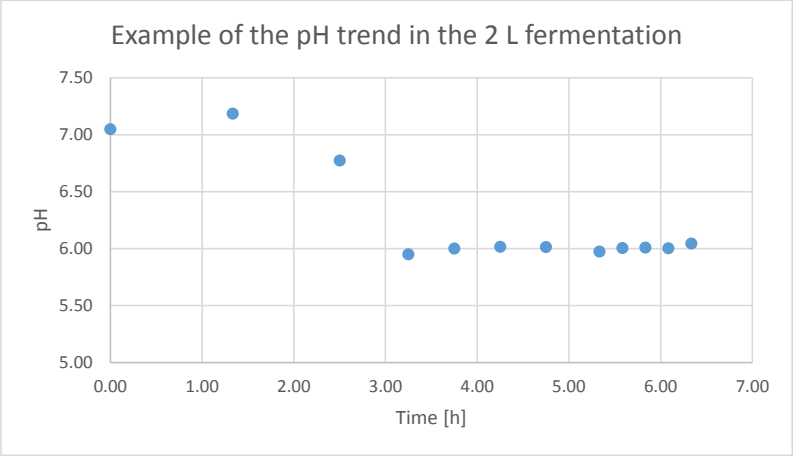


Figure S7. Velocity in the reactor domain for three different mesh sizes. Left: Velocity across the horizontal line in Fig. S6. Right: Velocity across the vertical line in Fig. S6.

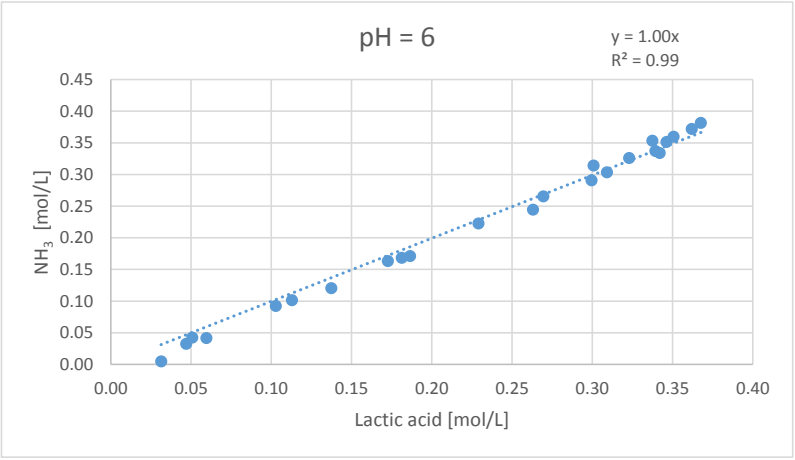
Fermentation no.	Data					
	Time [h]	NH3 24% [g]	NH3 conc	Lactic acid	Lactic acid	pH
	0.00	0.00	0.00	0.00	0.00	7.05
	1.33	0.00	0.00	0.00	0.00	7.18
	2.50	0.00	0.00	1.01	0.01	6.77
	3.25	0.70	0.00	2.82	0.03	5.95
	3.75	5.90	0.04	5.36	0.06	6.00
	4.25	13.10	0.09	9.27	0.10	6.02
	4.75	23.17	0.16	15.56	0.17	6.01
	5.33	37.70	0.27	24.27	0.27	5.97
	5.58	43.10	0.30	27.85	0.31	6.01
	5.83	46.28	0.33	29.09	0.32	6.01
	6.08	47.40	0.33	30.80	0.34	6.00
L8	6.33	47.90	0.34	30.57	0.34	6.04
	0.00	0.00	0.00	0.00	0.00	6.91
	2.45	0.00	0.00	2.05	0.02	6.10
	3.03	4.60	0.03	4.23	0.05	5.99
	3.70	14.40	0.10	10.17	0.11	6.04
	4.15	23.90	0.17	16.32	0.18	6.00
	4.45	31.60	0.22	20.63	0.23	6.00
	4.95	44.59	0.31	27.09	0.30	6.02
	5.37	50.20	0.35	30.40	0.34	6.02
L12	5.78	51.10	0.36	31.58	0.35	6.00
	0.00	0.00	0.00	0.00	0.00	6.99
	1.55	0.00	0.00	0.51	0.01	6.83
	2.55	0.20	0.00	2.10	0.02	6.13
	3.22	6.00	0.04	4.57	0.05	6.01
	4.05	17.09	0.12	12.38	0.14	6.00
	4.38	24.29	0.17	16.80	0.19	6.01
	4.80	34.70	0.24	23.70	0.26	6.00
	5.05	41.28	0.29	26.99	0.30	6.00
	5.53	49.90	0.35	31.19	0.35	6.00
	6.38	52.80	0.37	32.60	0.36	6.00
L14	7.13	54.20	0.38	33.12	0.37	5.99

L25	0.00	n.d.	0.00	0.00	6.83
	0.85	n.d.	0.00	0.00	6.82
	1.63	n.d.	0.30	0.00	6.74
	2.07	n.d.	0.62	0.01	6.59
	2.47	n.d.	1.28	0.01	6.31
	2.97	n.d.	2.67	0.03	5.95
	3.47	n.d.	5.36	0.06	5.97
	3.97	n.d.	9.67	0.11	5.96
	4.47	n.d.	16.32	0.18	6.01
	4.97	n.d.	23.68	0.26	5.98
	5.47	n.d.	29.64	0.33	5.90
	5.97	n.d.	31.43	0.35	5.92
	6.47	n.d.	31.77	0.35	5.92

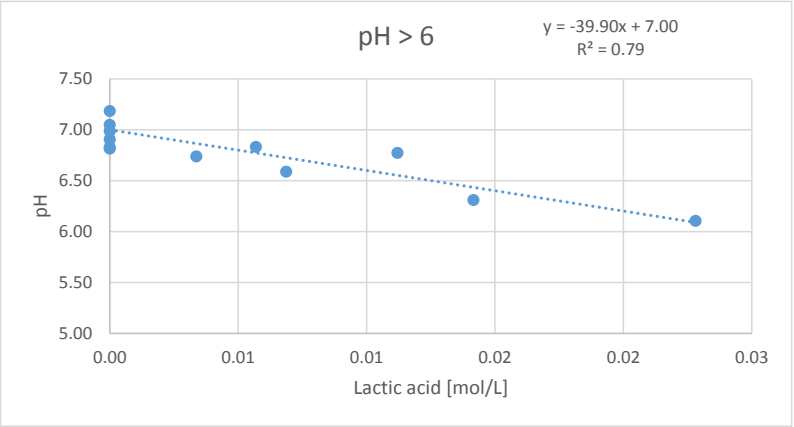
Volume of bioreactor [l]	NH3 [g/mol]	Lactic acid [g/mol]
2	17.04	90.08



The pH droped from pH ca. 7 to 6 in the first 3 hours of the fermentation. Then, the pH was controlled at pH = 6 by adding ammonia.
For the correlation between the pH, lactic acid and ammonia the data of four 2-L fermentations (see raw data on the left) was analyzed.
At pH = 6, the addition of ammonia was 1:1 [mol/L] proportional with the synthesis of lactic acid (see next graph).



For the beginning of the fermentation, when the pH drops, only a few data points were available. A linear correlation to predict the pH from the lactic acid concentration in this zone is shown in the next graph.



A linear correlation to predict the pH from the lactic acid and ammonia concentration was derived:

$$\mathrm{pH} = -39.9 \cdot (LA - 1 \cdot NH) + 7.00$$

Here the concentrations are given in mol/L.
Note that the concentrations in the publication were presented in g/L!

Paper F

A Validated CFD-based Compartment Model to Assess pH Gradients in Lactic Acid Bacteria Cultivations

Robert Spann[a], Christophe Roca[b], Krist V. Gernaey[a], Gürkan Sin[a]

[a] Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

[b] Chr. Hansen A/S

Submitted to the AIChE Journal (Online ISSN:1547-5905).

Abstract

A compartment model (CM) was designed based on a computational fluid dynamics (CFD) simulation, and applied to analyze the mixing effects on the cultivation performance of lactic acid bacteria in a 700 L bioreactor. To this end, the kinetic model, which describes the biomass growth as well as pH dynamics in the cultivation, was coupled with the CM describing mixing and transport phenomena within the bioreactor. A Monte Carlo simulation was performed to account for uncertainties in the model. Applying the validated CM, a better base addition position for pH control was proposed to reduce the pH gradients in the reactor. The computation speed of the CM (<2 s for one simulation) enables fast off-line process condition testing and on-line applications, e.g. as a soft sensor, in contrast to a CFD simulation that takes several hours/days to simulate. This allows future application of the CM for on-line monitoring and control.

Introduction

Biotechnological processes demand continuous improvement in order to be competitive. Mechanistic models are built and shown to be useful to support bioprocess development efforts at different levels; e.g. to support process understanding, to speed up the development of new processes, and to monitor and control the production process to ensure that the right product quality and quantity are achieved.[1,2] However, the cell physiology, and hence often the product quality are different at production scale compared to lab scale because of altered cultivation conditions.[3] In large-scale processes, culture parameters such as the carbon source concentration, oxygen concentration, and the pH among others are not homogeneously distributed in contrast to traditional lab-scale conditions. Transport limitations exist due to a lower mixing capability of large-scale bioreactors. For example, the addition of alkali solution to a batch cultivation in order to control the pH leads to a significantly higher pH in the vicinity of the inlet.[4] Cells are exposed to changing conditions while they circulate through the bioreactor, and they might adapt to those conditions resulting in increased maintenance requirements and a lower yield at large scale.[5–8] This makes it challenging for both the experimental process design that usually takes place at lab scale, and for the model development that relies typically on data derived from lab-scale experiments. They often do not consider the heterogeneous conditions at large scale.

Computational fluid dynamics (CFD) models are applied to investigate the mixing in bioreactors. They have been used to investigate potential improvement of the mass transfer in large-scale bioreactors, e.g. the oxygen transfer rate by optimizing the impeller configuration.[8,9] Furthermore, several process conditions during a cultivation have been investigated by coupling biokinetic models with the CFD model.[10–13] The drawback of CFD models is, however, that they require a long computation time of more than a day to simulate a full-scale cultivation[14] – which might change in future e.g. with quantum computing. Even though CFD simulations are a powerful tool for process development, CFD models are unsuitable for on-line applications today due to their high computational demand.

Compartment models therefore offer an opportunity as a compromise between the computational complexity for describing mixing and the simulation speed that is needed for various applications such as on-line monitoring and control applications, as well as fast off-line simulations, e.g. to test different scenarios for reactor geometry, mixing equipment and sensor locations. Like CFD models, they consider the mixing and the resulting heterogeneous process conditions in large-scale bioprocesses. In the compartment model, the spatial resolution is reduced compared to a CFD model and the flow rates between compartments are usually determined based on the velocity fields predicted by a steady state CFD simulation.[15] So far, various bioprocesses have been simulated with compartment models, in order to investigate

gradients of the substrate, oxygen, and metabolites.[15–19] However, pH gradients have not been investigated despite the importance of the pH on the microbial physiology.[20,21]

The objective of this study was therefore to design and validate a CFD-based compartment model for a *Streptococcus thermophilus* batch cultivation, and to predict the pH gradients that occurred in the 700 L bioreactor. To this end, first a compartment model was designed and the flow rates were extracted from a steady state CFD simulation validated in a previous study (Spann et al., 2018, submitted: "CFD predicted pH gradients in lactic acid bacteria cultivations"). The compartment model was validated with tracer experiments and the CFD predictions. Then, a biokinetic model describing microbial growth, and a chemical model describing the pH dynamics in the bioreactor were integrated in the compartment model, and a pH controlled batch cultivation was simulated. The model predictions were compared with experimental multi-position pH measurements. Finally, a further application of the compartment model for process design purposes was demonstrated, in which the compartment model was applied to test different alkali addition positions in order to decrease the pH gradients during the cultivation.

Materials and Methods

Cultivation Conditions and Off-Line Analysis

A *Streptococcus thermophilus* cultivation was performed in a stirred tank bioreactor (Chemap AG, Switzerland) with 700 L liquid volume in batch mode. The bioreactor was equipped with three 6-blade Rushton turbines and four baffles (see for the details Spann et al., 2018, submitted). The cultivation was performed with a stirrer speed of 130 rpm, at 40 °C, and with nitrogen headspace gassing. 24 % (w/v) ammonia solution (NH_4OH) was added below the bottom impeller to control the pH at the set point pH 6. The pH sensor that provided the measurements as input to the controller was located 35 cm above the bottom of the bioreactor close to the bioreactor wall. The pH was furthermore monitored every second at heights of 0.10 m and 1.60 m with CPS471D pH sensors (Endress+Hauser AG, Switzerland). The medium contained 70 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$.

The dry cell weight was determined from cultivation broth samples that were first centrifuged, then washed twice with 0.9 % NaCl, and finally dried at 70 °C for 24 h. Organic acids and sugars were quantified from filtered (0.2 µm) samples in an HPLC (Dionex UltiMate 3000, Thermo Fisher Scientific, Waltham, MA). It was equipped with a refractive index detector (ERC RefractoMax 520) and an Aminex® HPX-87H column (Bio-Rad Laboratories, Hercules, CA), and operated at a flow rate of 0.6 mL min^{-1} at 50 °C using 5 mM H_2SO_4 according to suppliers instructions.

Compartment Model

A compartment model was designed based on the steady state velocity profiles that were obtained from a CFD simulation. The CFD simulations that were conducted in ANSYS CFX 17.1 (ANSYS, Inc., US-PA) were described elsewhere (Spann et al., 2018, submitted). The CFD model captured half of the bioreactor volume applying a rotational periodicity plane. The mesh was defined with 6-sided hexahedral elements and consisted of $1.6 \cdot 10^6$ nodes in total in both the stationary domain and the rotating domains. Steady state velocity profiles were obtained representing a stirrer speed of 130 and 240 rpm using the k-ε turbulence model. The CFD model was validated at 240 rpm with NaOH tracer pulse experiments from the top of the bioreactor using multi-position pH measurements. Both the dynamic pH change and the mixing time have been accurately predicted by the CFD model.

To design the compartments for the reactor model, it was decided to represent each recirculation loop that was revealed by the velocity profiles of the CFD simulation as one compartment (Figure 1 and Supplementary Figure S1). To further define the boundaries and the inflow and outflow of fluids, the following procedure was used: (i) Horizontal planes were set up every 1 cm in CFX-Post and the axial velocities and node areas were exported for each plane from the CFD results. (ii) The flow rates were calculated as the product of the velocity and the area, and the positive and negative axial flow rates were separated. (iii) The arithmetic means of both the positive and negative flows were calculated. (iv) The local minima (here five) of the mean positive axial flow rates defined the interfaces between the compartments (here six compartments) (Supplementary Figure S2). These positions matched with the local maxima of the mean negative flow rates over the bioreactor height. (v) The arithmetic mean of the mean positive flow and of the absolute mean negative flow of the interface planes were set as the flow between the compartments. In this way, a continuity is ensured, which avoids mass accumulation in compartments. The properties of the compartment model can be found in Table I.

In addition, a 7th compartment was designed capturing the 10 cm at the top of the bioreactor that was necessary to support the tracer pulse simulation (see below). The resulting configuration of the compartmental model is shown in Figure 1 B.

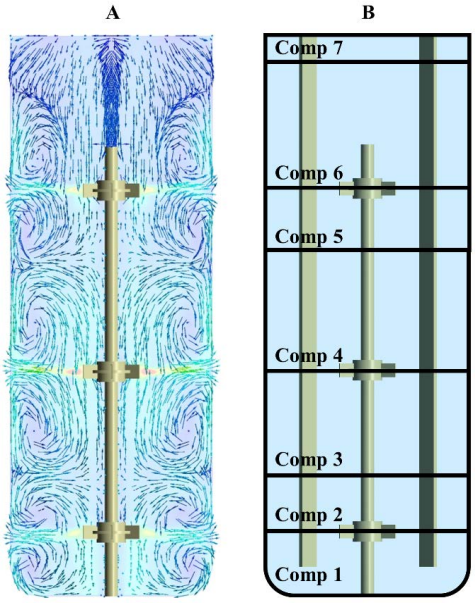


FIGURE 1. Design of the CFD- based compartment model for a stirred speed of 130 rpm. Velocity streamlines of the steady state CFD solution (A). The seven compartments for the 700 L bioreactor (B).

The compartment model has been implemented in MATLAB (The MathWorks®, Natick, MA) as an ordinary differential equation (ODE) system. To simulate the lactic acid bacteria cultivation the biokinetic and the pH model (see below) were defined together with the compartment model that represented the stirrer speed of 130 rpm in the ODE system.

For comparison, a one-compartment model with a volume of 700 L was simulated in MATLAB to model the cultivation without the effects of gradients.

Table I. Properties of the compartment models

Compartment	Volume [L]		Compartment interconnection	Flow rates between the compartments [L s^{-1}]	
	130 rpm	240 rpm		130 rpm	240 rpm
1	38.8	38.8	1 ↔ 2	12.2	17.9
2	36.4	52.6	2 ↔ 3	8.8	12.1
3	67.1	50.9	3 ↔ 4	14.2	18.1
4	76.8	60.6	4 ↔ 5	9.7	13.6
5	39.6	55.7	5 ↔ 6	11.7	17.5
6	80.6	80.6	6 ↔ 7	10	15.3
7	19.1	19.1			

Mixing Time Calculation with a Tracer Pulse

The addition of a tracer pulse was simulated by starting a simulation with 100 g L^{-1} of an additional state variable in the top compartment while the additional variable was not present in the other compartments. Only the flows between the compartments were modelled whereas the reaction rates of the dynamic model (see below) were not considered. The concentration of the additional variable in each compartment was then normalized (Eq. (1)) and the mixing time was calculated with the logarithmic squared deviation with respect to the normalized upper bound 1 (Eq. (2)).[22] 95 % homogeneity was achieved at $\log D^2 = -2.6$.

$$C'_{tracer,i}(t) = \frac{C_{tracer,i}(t) - C_{tracer,i}(t=0)}{C_{tracer,i}(t=5\ \text{min}) - C_{tracer,i}(t=0)} \quad (1)$$

where $C'_{tracer,i}$ represents the normalized tracer concentration in the i-th compartment, and $C_{tracer,i}(t)$ the tracer concentration in the i-th compartment at time t after the pulse.

$$log\ D^2 = log\left[\frac{1}{n} \cdot \sum_{i=1}^{n} \left(C'_{tracer,i}(t) - 1\right)^2\right] \quad (2)$$

where n is the number of compartments. In this study, the 5 lower compartments were considered to calculate the mixing time, as this represented the experimental conditions of the tracer pulses (Spann et al., 2018, submitted). In these experiments, the bioreactor was filled with 700 L 35 °C tap water, and pulses of a NaOH solution (27 %, Novadan ApS, Denmark) were applied from the top of the bioreactor within a pH range of 5.0 to 6.0. Five pH sensors (CPS471D) measured the pH every second at different positions, namely at heights of 0.10, 0.35, 0.60, 0.95, and 1.25 m with respect to the bottom of the bioreactor with a clearance of 0.1 m to the bioreactor wall.

Biokinetic and pH Model

A dynamic mechanistic model was implemented that described the evolution of the state variables in the *S. thermophilus* cultivation. The model had been developed and validated earlier with 2 L experiments and is described in detail in Spann et al.[23] The model comprises a biological model part and a chemical model part.

The biokinetic model

The biological model predicted the biomass growth, substrate (lactose) consumption, and the lactic acid secretion by the bacteria among others. The biokinetic model is based on the overall process stoichiometry[24] (Eq. (3)-(4)). The biomass growth rate is modelled as a product of functions that account for the lag-time (f_{lag}), lactose inhibition and limitation (f_S)[25], lactate inhibition (f_P)[26], and the pH in the cultivation broth (f_{pH})[27] (Eq.(5)-(6)).

$$Lactose + Ammonia + Phosphoric\ acid \rightarrow Biomass + Lactic\ acid + Galactose \quad (3)$$

$$q_S \cdot CH_2O + q_{NH} \cdot NH_3 + q_{Ph} \cdot H_3PO_4 \rightarrow q_X \cdot CH_aO_bN_cP_d + q_P \cdot CH_2O + q_{Gal} \cdot CH_2O \quad (4)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot f_{lag} \cdot f_S \cdot f_P \cdot f_{pH} \cdot C_X \quad (5)$$

$$\frac{dC_X}{dt} = \mu_{max} \cdot \left(1 - e^{-\frac{t}{t_{lag}}}\right) \cdot \frac{C_S}{C_S + K_s + \frac{C_S^2}{K_I}} \cdot \frac{1}{1 + e^{K_{P,La}(C_{LA} - K_{La1})}} \cdot e^{-\left(\frac{(pH_{opt} - pH)^2}{\sigma_{pH}^2}\right)} \cdot C_X \quad (6)$$

with K_{La1} dependent on the pH:

$$K_{La1} = K_{La} \cdot \frac{1}{1 + e^{K_{P,pH1} * (pH - K_{P,pH2})}} \quad (7)$$

The lactic acid secretion rate of the bacteria is modelled as a growth dependent function:[28]

$$\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt} \quad (8)$$

The lactose uptake rate is the sum of the biomass growth and lactic acid secretion rate (Eq. (9)). Y_{gal} accounts for the secretion of galactose since the used strain does not metabolize the galactose that is produced from the lactose.

$$\frac{dC_S}{dt} = -\left(1 + Y_{gal}\right) \cdot \left(\frac{dC_X}{dt} + \frac{dC_P}{dt}\right) \quad (9)$$

The kinetic parameters (Table II) were identified in a parameter estimation that included an identifiability and uncertainty analysis.[23]

Table II. Kinetic parameters for the *S. thermophilus* model

Symbol	Value		Std. deviation	Reference
	CFD model	Monte Carlo model		
Biological model				
K_I	164 g L^{-1}	164 g L^{-1}	n.d.	[25]
K_{La}	21.1 g L^{-1}	19.80 g L^{-1}	0.05 g L^{-1}	[23]
$K_{P,La}$	0.2 L g^{-1}	0.24 L g^{-1}	0.03 L g^{-1}	[23]
$K_{P,pH1}$	20	20	n.d.	[23]
$K_{P,pH2}$	7	7	n.d.	[23]
K_S	0.79 g L^{-1}	0.79 g L^{-1}	n.d.	[25]
pH_{opt}	6.22	6.39	0.06	[23]
t_{lag}	0.38	0.38	n.d.	parameter estimation
Y_{gal}	0.63 g g^{-1}	0.69 g g^{-1}	0.04 g g^{-1}	[23]
α	5.59 g g^{-1}	5.19 g g^{-1}	0.01 g g^{-1}	[23]
μ_{max}	2.16 h^{-1}			[23]
σ_{pH}	1.09	1.42	0.04	[23]
Mixed weak acid/base model				
$K'_{r,C1}$	10^7 s^{-1}			[29]
$K'_{r,LA}$	10^7 s^{-1}			[29]
$K'_{r,NH}$	10^{12} s^{-1}			[29]
$K'_{r,P1}$	10^8 s^{-1}			[29]
$K'_{r,P2}$	10^{12} s^{-1}			[29]
$K'_{r,W}$	10^{10} s^{-1}			[29]
$K'_{r,Z}$	10^7 s^{-1}			[29]
pK_{C1}	$3404.7/(T - 14.8435 + 0.03279 \cdot T)$			[30]
pK_{LA}	3.86			[31]
pK_{NH}	$2835.8/(T - 0.6322 + 0.00123 \cdot T)$			[30]
pK_{P1}	$799.3/(T - 4.5535 + 0.01349 \cdot T)$			[30]
pK_{P2}	$1979.5/(T - 5.3541 + 0.01984 \cdot T)$			[30]
pK_W	14			[30]
pK_Z	9.4			[23]
T	313 K			measured process condition
Initial Conditions				
$C_{Gal,t=0}$	0.0 g L^{-1}			
$C_{Glc,t=0}$	0.0 g L^{-1}			
$C_{P,t=0}$	0.0 g L^{-1}			
$C_{S,t=0}$	70 g L^{-1}		2.3 g L^{-1}	
$C_{tCO,t=0}$	$1.002 \cdot 10^{-5}$ $mol\ L^{-1}$			
$C_{tNH,t=0}$	0.005 g L^{-1}			
$C_{tPh,t=0}$	2 g L^{-1}			
$C_{tZ,t=0}$	2 mol L^{-1}			
$C_{X,t=0}$	0.025 g L^{-1}		$8 \cdot 10^{-4}$ g L^{-1}	

The chemical model to predict the pH

The objective of the chemical model was to predict the pH in the cultivation broth. Two models were applied to predict the pH, namely a linear equation and a mixed weak acid/base model.

In this study, first the linear equation that had been used in the CFD study (Spann et al., 2018, submitted) was applied to predict the pH (Eq. (10)). It therefore allowed a comparison of the CFD and compartment model predictions but is dependent on the data it has been derived from. The mixed weak acid/base model (see below) could not be solved in the CFD simulation because the fast reaction rates resulted in a stiff system of differential equations.

$$pH = -0.44 \cdot \left(C_P - 5.29 \cdot C_{NH_3}\right) + 7.00 \tag{10}$$

Second, a mechanistic mixed weak acid/base model was applied that had been previously validated for the *S. thermophilus* cultivation.[23] It comprises the dissociation reactions of lactic acid, ammonia, phosphoric acid, carbonic acid, water, and an unspecified compound Z (Table III).[29] The mixed weak acid/base model could not be solved in the CFD simulation because the fast reaction rates resulted in a stiff system of differential equations.

Table III. Kinetics of the mixed weak acid/base model

Dissociation reaction	Reaction rate vector
$C_3H_6O_3 \leftrightarrow C_3H_5O_3^- + H^+$	$\mathrm{K'_{r,LA}} \cdot \mathrm{K'_{LA}} \cdot [C_3H_6O_3] - K'_{r,LA} \cdot [C_3H_5O_3^-] \cdot [H^+]$
$NH_4^+ \leftrightarrow NH_3 + H^+$	$\mathrm{K'_{r,NH}} \cdot \mathrm{K'_{NH}} \cdot [NH_4^+] - K'_{r,NH} \cdot [NH_3] \cdot [H^+]$
$H_3PO_4 \leftrightarrow H_2PO_4^- + H^+$	$\mathrm{K'_{r,P1}} \cdot \mathrm{K'_{P1}} \cdot [H_3PO_4] - K'_{r,P1} \cdot [H_2PO_4^-] \cdot [H^+]$
$H_2PO_4^- \leftrightarrow HPO_4^{2-} + H^+$	$\mathrm{K'_{r,P2}} \cdot \mathrm{K'_{P2}} \cdot [H_2PO_4^-] - K'_{r,P2} \cdot [HPO_4^{2-}] \cdot [H^+]$
$H_2CO_3^* \leftrightarrow HCO_3^- + H^+$	$\mathrm{K'_{r,C1}} \cdot \mathrm{K'_{C1}} \cdot [H_2CO_3^*] - K'_{r,C1} \cdot [HCO_3^-] \cdot [H^+]$
$H_2O \leftrightarrow OH^- + H^+$	$\mathrm{K'_{r,W}} \cdot \mathrm{K'_{W}} - K'_{r,W} \cdot [OH^-] \cdot [H^+]$
$ZH^+ \leftrightarrow Z + H^+$	$\mathrm{K'_{r,Z}} \cdot \mathrm{K'_{Z}} \cdot [ZH^+] - K'_{r,Z} \cdot [Z] \cdot [H^+]$

To account for the changing ionic strength in the cultivation broth due to the increase of the lactate and ammonium concentrations, the activity coefficients were first calculated by a modified Debye-Hückel model by Davies[32] (Eq. (11)) in order to then determine the apparent equilibrium constants $\mathrm{K'_{LA}}$, $\mathrm{K'_{P1}}$, $\mathrm{K'_{P2}}$, $\mathrm{K'_{C1}}$, $\mathrm{K'_{W}}$, and $\mathrm{K'_{Z}}$ according to Musvoto et al.[29]

$$log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3 \cdot I\right) \tag{11}$$

With the ionic strength (I) of the charged components (i):

$$I = \frac{1}{2}\sum_i z_i^2 C_i \tag{12}$$

The pH was controlled at the pH_{set} = 6.0 by adding ammonia solution using a P-controller with the controller gain (K_P) of 5 mol L^{-1} · liquid volume [L]:

$$NH_4OH_{add} = K_P \cdot (pH_{set} - pH) \tag{13}$$

Monte Carlo Simulation

A Monte Carlo simulation of the dynamic model was performed to account for uncertainties in the model. The Monte Carlo procedure consisted of three steps: First, the input uncertainties were identified and defined. In this case study, uncertainties in the kinetic model parameters and the initial substrate and biomass conditions were considered. The range of uncertainties are listed in Table II as defined earlier.[23] Second, N = 200 random input samples were generated from the input uncertainty domain using the Latin hypercube sampling technique.[33,34] The sampling matrix can be found in the Supplementary Figure S3. Third, the Monte Carlo simulation was performed. The model predicted therefore a probability distribution of the model outputs.

Results and Discussion

The compartment model was designed with the aim to fulfill two key characteristics: (i) The model should accurately account for the fluid dynamics and thus heterogeneous process conditions in the large scale bioreactor. (ii) It was meant to reduce the computation time considerably compared to a CFD simulation. A fast computation time was crucial, on the one hand, for a widespread use in different process development tasks, and, on the other hand, to be suitable for on-line monitoring and control applications. To achieve this, a manual CFD-based compartmentalization was performed where each of the recirculation loops that have been predicted by the CFD simulation were considered as a compartment. The seventh compartment was placed for the top 10 cm of the bioreactor that enabled the tracer pulse simulation in order to validate the compartment model. Two compartment models were designed: The first model modelled the stirrer speed of 130 rpm, which corresponded to the conditions of the *S. thermophilus* cultivation. The second model modelled the stirrer speed of 240 rpm, which corresponded to the conditions for the tracer pulse experiments.

Validation of the Compartment Model with the Mixing Time

The compartment models were validated with a tracer pulse simulation from the top of the bioreactor. The predictions of the CFD model were used as a benchmark to assess the compartment model since the CFD model was dynamically validated with experimental measurements (Spann et al., 2018, submitted). It had been shown that the CFD model is capable of predicting both the dynamic trend of the tracer concentration at various positions in the bioreactor after a tracer pulse, and to predict the mixing time accurately.

The mixing times to achieve 95 % homogeneity were compared with the experiments and the CFD model (Table IV). The compartment model that mimicked a stirrer speed of 240 rpm predicted 52 s to achieve 95 % homogeneity, which were both in good agreement with the experimentally measured mixing time values (48 ±5 s). In addition, the mixing time simulated by the compartment model matched well with the mixing time prediction of the CFD simulation, which was 55 s. A further comparison of the mixing time prediction was made at 130 rpm (the impeller speed of the cultivation), which showed that the prediction of the compartment model was 68 s compared to the 89 s mixing time predicted by the CFD model.

Table IV. Mixing times that were predicted by the compartment model, the steady state CFD model, and the experiments

	Mixing time [s] to reach 95% homogeneity	
	130 rpm	240 rpm
Experiment	not available	48 ± 5
CFD model	89	55
Compartment model	68	52

In general, the predicted mixing times were shorter with the compartment models than the CFD models, which may be explained by the reduced resolution of the compartment model. When the tracer enters the next compartment, the same concentration is assumed in the whole compartment, while many more mesh elements must be passed in the CFD model (the complete CFD model consists of ca. 1.6 million nodes). The present compartment model was coarse as only seven compartments were designed. Axial (up and down) mixing within the recirculation loops was therefore not represented in the compartments. Horizontal flows (radial and circumferential) have not been considered in the presented compartment model. Nevertheless, for example, the circumferential flow might be faster close to the impellers than at the bioreactor wall and the baffles.

The simple compartment model with seven compartments was not expected to describe the fluid dynamics comprehensively (e.g. as comprehensive as CFD). Nonetheless, the results validated the compartment model against both measurements and CFD predictions to describe mixing adequately in the reactor for this case. Indeed, the relative difference between the prediction of CFD and the compartmental model for the mixing time were lower than 20% at 130 rpm. This difference was only 5% at 240 rpm. Anyhow, as the growth rate constants of microbial system is slower compared to the mixing time constants, this mixing time accuracy was deemed acceptable for this study. These results are specific to the reactor geometry as well as cultivation conditions (viscosity etc.) used in this study. If the conditions in the bioreactor are modified, e.g. a lower or higher stirrer speed, the compartment model properties (compare Table I) might change. It is noted that a more complicated compartment model design might be needed to adequately describe mixing in other systems. For instance, the study of Delafosse et al. required more than 12000 compartments to simulate the expected mixing time.[35] It was not the aim of this study to optimize the compartment model design further.

Simulation of the Lactic Acid Bacteria Cultivation

In this section, the validated compartment model describing the fluid dynamics, was combined with the biokinetic and pH model in order to simulate the lactic acid bacteria cultivation. The results were compared with experimental measurements, the CFD model, and a homogeneous model. The cultivation was performed at 130 rpm and hence the corresponding compartment model was used. The same model equations (including the pH calculation (Eq. (10)) and parameters were used as in the CFD simulation (Spann et al., 2018, submitted) to allow a comparison of the model predictions.

The *S. thermophilus* batch cultivation was simulated and the prediction of the state variables biomass, lactic acid, and lactose was assessed (Figure 2). The three models, the compartment model (dashed line), the CFD model (solid line), and a one-compartment model (dotted line) match the off-line measurements (circles) quantitatively. The off-line measurements are shown

for comparison purposes here. The objective of the comparison was to evaluate the effect of different modelling approaches for the fluid dynamics combined with the same kinetic parameter set. The difference between the compartment model and the CFD model was less than 0.42 g L^{-1} in the biomass prediction (Figure 2 A), < 0.97 g L^{-1} in the lactic acid prediction (Figure 2 B), and < 3.74 g L^{-1} for the lactose concentration (Figure 2 B). This corresponds to a difference of less than 10 % with respect to the maximum concentrations that were recorded. The difference between the compartment model and the homogeneous model was considerably lower. It was less than 0.15 g L^{-1} in the biomass prediction (Figure 2 A), < 0.77 g L^{-1} in the lactic acid prediction (Figure 2 B), and < 1.56 g L^{-1} for the lactose concentration (Figure 2 B), corresponding to a difference of less than 3 %.

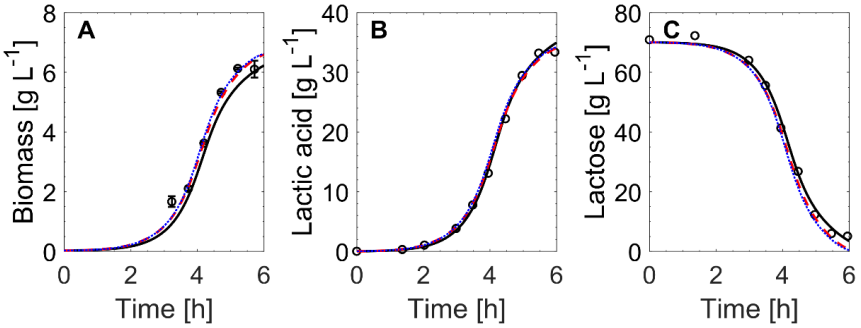


Figure 2. Simulated and measured biomass (A), lactic acid (lactate + lactic acid) (B), and lactose (C) concentrations in the 700 L *S. thermophilus* cultivation. Measurements (open circles), CFD simulation (black solid line), compartment model (red dashed line), homogeneous simulation (blue dotted line).

As the difference between the homogeneous model and the compartment model was small, it might be argued that the compartment model is not needed. However, the purpose of the compartment model is to be able to describe the effects of the mixing, e.g. the pH gradients. What if scenarios for different location of acid/base addition for pH control are therefore possible, which would otherwise not be possible with the homogeneous model assuming a completely mixed bioreactor. Moreover, the results are case specific and hold only true for the investigated strain and process conditions. More importantly, the comparison showed only the view of the total bioreactor, whereas local gradients, both of the pH and the state variables, might exist in the cultivation and were evaluated below.

The difference between the CFD model and the compartment model was visible and it could be attributed to the ODE solving algorithms or the simulated process conditions, among others. Discretization errors may have caused the difference since the CFD model was solved with a time step of 1 s (Spann et al., 2018, submitted), while the MATLAB ode15s solver uses a dynamic step size providing a more accurate integration of the stiff ODE system (the integration error was set to 10^{-5}). Furthermore, gradients of state variables, such as substrate, lactic acid, and pH were predicted differently, and could have led to divergent model results as they influence the biomass growth, e.g. by inhibition (see Eq. (6)). The CFD simulation predicts a similar gradient as the compartment model (see below) at the measurement positions, which were close to the bioreactor wall. However, in contrast to the compartment model, the resolution of the CFD simulation is finer, and hence the high pH in the close vicinity of the base addition affected the CFD results. Biomass growth is consequently reduced in the base

addition region in the CFD simulation, while an “average” pH is modelled in the bottom compartment of the compartment model. Substrate and lactic acid gradients are from a statistical point of view nonexistent (see below).

The computation time for the compartment model was less than 2 s on an Intel® Core™ i7-5600U CPU @2.6 GHz (1000 repetitions showed this performance), which was considerably faster than the 4 days on 20 CPU cores on the DTU High Performance Computing Cluster (https://www.hpc.dtu.dk/) that was required for the CFD simulation. The long computation time is generally the issue with simulations performed with biokinetic models coupled with CFD models.[14]

Monte Carlo simulation

The fast simulation time of the compartment model allowed considering a Monte Carlo simulation of the mechanistic mixed weak acid/base model together with the biokinetic model in the subsequent compartment model implementations. With the Monte Carlo simulation the uncertainties of the model parameters could be covered as they have been determined in the parameter estimation.[23] This allowed a more robust model prediction in comparison to the deterministic simulation used above. Furthermore, the mechanistic pH model (Table III) replaced the data driven pH correlation (Eq. (10)) since we generally prefer a mechanistic model to describe the system because of the advantages of mechanistic models, e.g. the extrapolation capability and flexibility.[2] The computation time of the compartment model remained below 2 s with the mixed weak acid/base model included.

A Monte Carlo simulation of the compartment model with 200 input samples was performed, and the mean concentration of the state variables over all compartments was evaluated (Figure 3). The assessment of the gradients between the compartments is shown in the subsequent section. The probabilistic compartment model predicted the state variables quantitatively, namely the biomass (Figure 3 A), lactic acid (Figure 3 B), and lactose (Figure 3 C) concentration. The Monte Carlo procedure was applied in order to propagate the input uncertainties (model parameters, and initial process conditions) to the model outputs.[36] The uncertainty in the model outputs is indicated by the 5 and 95 % percentiles.[37] The maximum standard deviation was $\sigma_{biomass} = 0.22$ g L^{-1}, $\sigma_{lactic\_acid} = 1.15$ g L^{-1}, and $\sigma_{lactose} = 4.15$ g L^{-1}. The uncertainties were deemed to be low and acceptable since they were in the range of expected measurement errors (for biomass and lactic acid) and of the anticipated medium preparation uncertainties (for lactose).

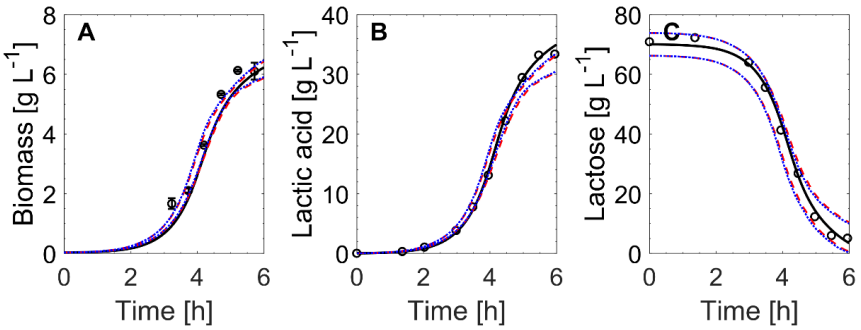


Figure 3. Simulated and measured biomass (A), lactic acid (lactate + lactic acid) (B), and lactose (C) concentrations in the 700 L *S. thermophilus* cultivation. Measurements with std. deviation (open circles), deterministic CFD simulation (black solid line), 5 and 95 % percentile of the Monte Carlo simulation output of the compartment model predictions (red dashed line), and the homogeneous model (blue dotted line).

Gradients in the Cultivation

The compartment model was applied to predict both the pH gradient and gradients of the biological state variables during the lactic acid bacteria cultivation.

pH gradients

The initial pH of the cultivation was pH = 6.8 that dropped due to lactic acid secretion within the first 2.5 h (Figure 4 A). The pH was then controlled to maintain it at pH = 6.0 by adding ammonia solution at the bottom of the bioreactor. A pH gradient was formed, which could be measured and predicted by the compartment model. The measurements showed the largest pH gradient at 4 h (during the exponential growth phase), with pH = 5.9 at the top of the bioreactor and pH = 6.2 at the bottom of the bioreactor. The pH measurements were taken every 1 s, and a few measurements showed a pH up to pH = 7.0. The mixed weak acid/base model predicted the initial pH drop accurately and the maximum pH gradient at 4 h to be between pH = 5.8 and 6.1 (Figure 4 A). The pH was raised in the bottom of the bioreactor because of the ammonia addition. Due to the fast lactic acid secretion and the insufficient mixing the pH decreased in the top compartments of the bioreactor. When the linear pH correlation was implemented in the compartment model, it matched the results of the mixed weak acid model closely (Figure 4 B).

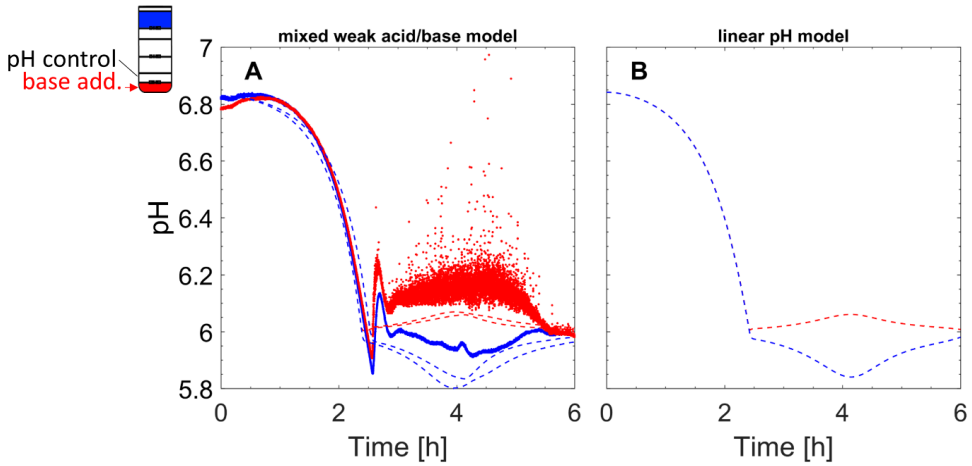


Figure 4. Compartment model predictions and measurements of the pH gradient in the 700 L *S. thermophilus* cultivation with a stirrer speed of 130 rpm. A: Monte Carlo simulation of the compartment model with the mixed weak acid/base model (dashed line) and measurements (dots) that were recorded every 1 s. B: Compartment model with the linear pH correlation.

The pH gradients were qualitatively predicted with a deviation of 0.1 pH units with the mixed weak mixed acid/base model. The small mismatch of the pH gradient predictions is likely related to the mixed weak acid/base model because it did not represent the medium components completely. The accuracy of the predictions could be improved in several ways: (1) Additional components such as acetic acid and amino acids that are present in the cultivation could be added to the mixed weak acid/base model. (2) The deviation of 0.1 pH units could also have resulted from calibration and measurement errors (e.g. drift of the pH sensors), which can be improved by the use of several pH sensors to double-check the results. (3) A finer compartment model design (i.e. increasing the number of compartments) can help better quantitatively represent the fluid dynamics (approximating that of the CFD simulation). However, this would lead to longer mixing times in the compartment model, and hence the predicted pH gradient might increase.

Nonetheless, the difference of 0.1 pH unit between the prediction of the compartmental model and the measurements is deemed acceptable, as this difference is not likely to cause significant errors on the microbial growth kinetics, hence this qualitative prediction of pH is deemed acceptable for the process design and monitoring purposes.

Gradients in the biological state variables

Apart from the pH gradients, differences of the concentrations of the biological state variables such as biomass, lactic acid, and lactose concentration were evaluated along the compartments. Figure 5 shows the differences at 4 h when the largest differences were predicted since the highest growth rate had been reached at that time. The predicted differences are statistically not significant, as the predicted deviation by the Monte Carlo simulation (Figure 5 top row) was much larger than the shown gradient of the mean values (Figure 5 bottom row). The biomass and lactic acid concentrations were slightly higher at the bottom than at the top of the bioreactor (Figure 5 A.2 and B.2). The lactose concentration was higher at the top of the

bioreactor than at the bottom (Figure 5 C.2). The differences were mainly attributed to the pH in the compartments. Owing to the pH function (f_{pH}) with the pH optimum at pH_{opt} = 6.4 (Eq. (6)), the growth rate was faster in the bottom compartments since the f_{pH} was higher there than in the top compartment where the pH was lower. Slightly more biomass and lactic acid were consequently accumulated in the bottom compartment, and more lactose was consumed there.

Such small differences in the lactic acid, and lactose concentration will not have an effect on the microbial performance. The situation could, of course, change if a fed-batch process would be applied since then a larger gradient of the carbon source would be expected as well.[6]

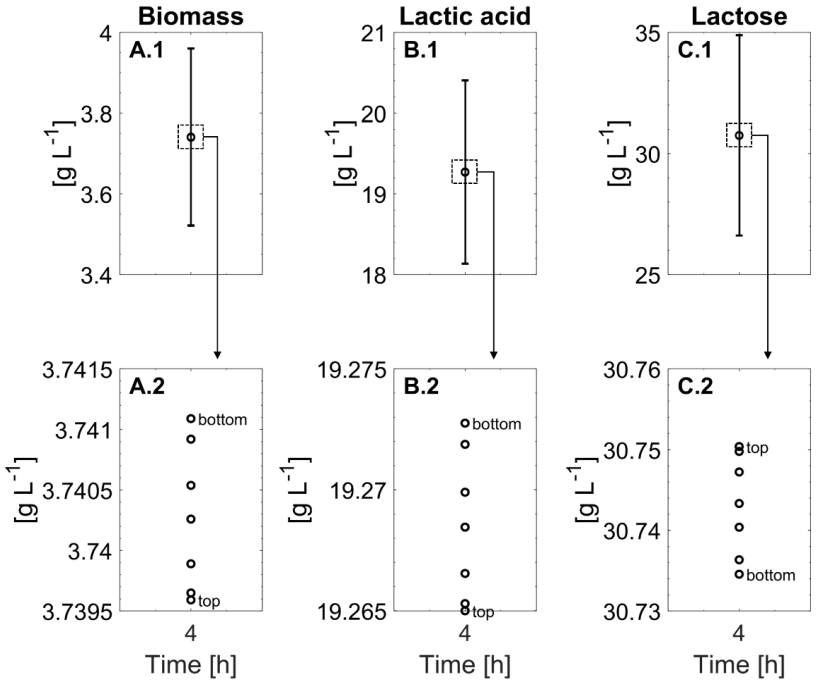


Figure 5. Biomass (A), lactic acid (lactate + lactic acid) (B), and lactose (C) concentrations gradients at 4 h of the *S. thermophilus* cultivation that were predicted by the compartment model. Std. deviation of the Monte Carlo simulation (top row). Mean concentrations in the seven compartments (bottom row).

The compartment model predicted the pH gradients qualitatively with a deviation of less than 0.1 pH units. The accuracy of the compartment model was therefore comparable with the CFD model (see Spann et al., 2018, submitted) whereas the computation time of the compartment model was much faster. The latter makes the compartment model a promising tool for on-line applications as it could be implemented in a soft sensor that requires a fast simulation time. The soft sensor could be applied for on-line risk-based monitoring and control, and could support the operation of the cultivation (manuscript in preparation).

Assessment of Different Base Addition Positions

The compartment model could be also applied for the process design besides the above mentioned on-line monitoring and control application. It could be applied to simulate different

scenarios, e.g. to test different control strategies or to investigate different base addition and pH sensor locations in order to decrease the pH gradient.

In the studied system, the base was added below the bottom impeller and the controlled pH was measured in compartment 2. Here, a pH gradient between 5.8 and 6.1 was predicted in the exponential growth phase as discussed above (Figure 4). In case the base addition would be placed underneath the middle impeller in the bioreactor, a pH gradient between 5.9 and 6.05 was predicted (Figure 6 A). If the base addition was placed below the top impeller while the controlled pH was still measured in compartment 2, a pH gradient between 5.95 and 6.3 was predicted (Figure 6 B). If the base would be added to the top of the bioreactor, a pH gradient between 5.95 and 7.0 was predicted (Figure 6 C).

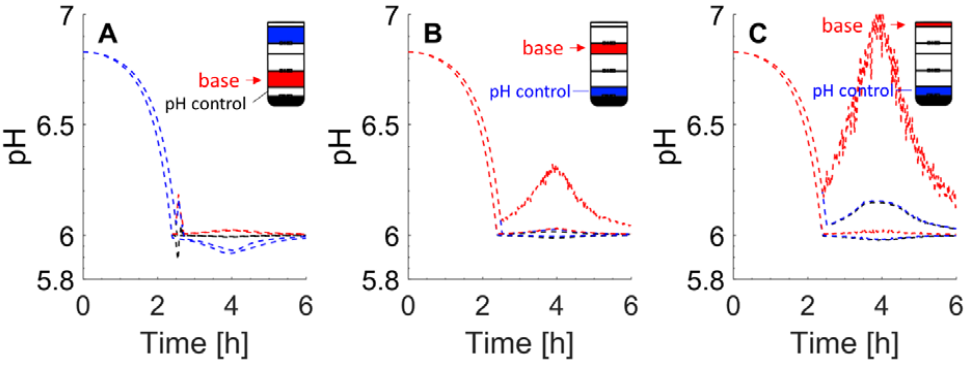


Figure 6. pH gradients predicted by the compartment model when the base would be added at different positions. Base addition below the middle impeller (A), below the top impeller (B), and from the top of the bioreactor (C).

According to these results, the pH gradient could be significantly reduced if the base inlet would be placed below the middle impeller. In the worst-case scenario, with the base addition from the top together with the measurement of the pH at the bottom (the measurement input to the controller here in compartment 2, Figure 6 C) the pH gradient would increase drastically. In general, other combinations of the position of both the base addition and the controlling pH measurement could be assessed. Experimental validation of the proposed design for the sensor location using the simulation of the compartmental model iteratively would help better improve optimization of the process.

Indeed, the compartment model could complement the process design phase at an early stage; in particular, it allows exploring and testing different scenarios with a short simulation time. Using a CFD simulation instead requires much longer simulation times and more computational resources.[19] However, we believe that a detailed analysis with a high resolution CFD simulation would be needed subsequently to substantiate the results. As an example, the gradients in the area in the vicinity of the base addition point could not be simulated with the compartment model. A higher pH is expected here as the base concentration is very high[4], which could be predicted by the CFD simulation with a higher spatial resolution (Spann et al., 2018, submitted).

Moreover, thanks to the promising results obtained in this study with a 700 L bioreactor, it is now intended to apply the CFD-based compartment model to larger (production-scale) bioreactors, e.g. > 50 m^3, and to support industrial production processes. The presented tools

can for example be applied to investigate to which extent pH gradients exist at production scale, and whether they could have an influence on the metabolic activity and especially the biomass growth and product quality.

Conclusion

A CFD-based compartment model was designed in order to represent the fluid dynamics of the 700 L bioreactor in an adequate manner and to reduce the computation time in comparison to CFD simulations. The compartment model was needed to account for the heterogeneous process conditions that occur in large-scale bioprocesses. It predicted the *S. thermophilus* batch cultivation including the pH gradients matching the CFD simulation and measurements with acceptable deviations. These gradients could have an influence on the microbial performance in the production process, while differences in the lactose and lactic acid concentrations along the height of the bioreactor were negligibly small. Thanks to the fast computation time without reducing the quality of the prediction significantly, the compartment model could be used for on-line applications, e.g. in a soft sensor for future risk-based monitoring and control. The computation speed also enabled a Monte Carlo simulation considering uncertainties in the model parameters and initial process conditions, which provided probabilistic model predictions in contrast to the deterministic CFD model prediction. Furthermore, the capability of the CFD-based compartment model in the early stage of the bioprocess design phase was demonstrated as different base addition positions were tested with the aim to minimize the pH gradients during the cultivation.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project). We are thankful for the discussion with Anna Eliasson Lantz (Technical University of Denmark, PILOT PLANT) and the cooperation with Chr. Hansen A/S.

Notation

C_{Gal}	galactose concentration (g L^{-1})
C_{Glc}	glucose concentration (g L^{-1})
C_{H+}	H^+ concentration (mol L^{-1})
C_{LA}	lactate concentration (g L^{-1})
C_{OH^-}	OH^- concentration (mol L^{-1})
C_P	total lactic acid (lactate and lactic acid) concentration (g L^{-1})
C_S	lactose (substrate) concentration (g L^{-1})
C_{tCO}	total carbonic acid ($H_2CO_3^*$ and HCO_3^-) concentration (mol L^{-1})
C_{tNH}	total concentration of NH_4^+ and NH_3 (g L^{-1})
C_{tPh}	total concentration of H_3PO_4, $H_2PO_4^-$, and HPO_4^{2-}(g L^{-1})
$C_{tracer,i}$	tracer concentration in the i-th compartment
$C'_{tracer,i}$	normalized tracer concentration in the i-th compartment
C_{tZ}	total concentration of the unknown compound (dissociated and undissociated form) (mol L^{-1})
C_X	biomass concentration (g L^{-1})
f_d	divalent activity coefficients (-)
f_{lag}	lag-time function (-)
f_m	monovalent activity coefficients (-)
f_P	lactic acid inhibition function (-)
f_{pH}	pH dependency function (-)
f_S	substrate limitation and inhibition function (-)
$H_2CO_3^*$	dissolved CO_2 and H_2CO_3
I	ionic strength (g L^{-1})
K'_{C1}	apparent equilibrium constant for the carbonic acid system (-)
K_I	substrate inhibition parameter (g L^{-1})
K_{La}	lactate inhibition parameter (g L^{-1})
K_{La1}	pH dependent lactate inhibition parameter (g L^{-1})
K'_{LA}	apparent equilibrium constant for the lactic acid system (-)
K'_{NH}	apparent equilibrium constant for the ammonia system (-)
K_P	P-controller controller gain
$K_{P,La}$	2. lactate inhibition parameter (L g^{-1})
$K_{P,pH1}$	lactate inhibition pH parameter (-)
$K_{P,pH2}$	2. lactate inhibition pH parameter (-)
K'_{P1}	apparent equilibrium constant for the phosphoric acid system (-)
K'_{P2}	apparent equilibrium constant for the dihydrogen phosphate system (-)
$K'_{r,C1}$	apparent reverse rate constant for carbonic acid dissociation (s^{-1})
$K'_{r,LA}$	apparent reverse rate constant for lactic acid dissociation (s^{-1})
$K'_{r,NH}$	apparent reverse rate constant for NH_4 dissociation (s^{-1})
$K'_{r,P1}$	apparent reverse rate constant for H_3PO_4 dissociation (s^{-1})
$K'_{r,P2}$	apparent reverse rate constant for $H_2PO_4^-$ dissociation (s^{-1})
$K'_{r,W}$	apparent reverse rate constant for water dissociation (s^{-1})
$K'_{r,Z}$	apparent reverse rate constant for the dissociation of the unknown component (s^{-1})
K_S	substrate limitation parameter (g L^{-1})
K'_W	apparent equilibrium constant for the water system (-)
K'_Z	apparent equilibrium constant for the unspecified compound system (-)
pH	pH of the cultivation broth
pH_{opt}	optimal pH parameter in the pH function (-)
pH_{set}	pH control set point (-)
pK_{C1}	pK_a constant for carbonic acid dissociation
pK_{LA}	pK_a constant for lactic acid dissociation
pK_{NH}	pK_a constant for NH_4 dissociation
pK_{P1}	pK_a constant for H_3PO_4 dissociation
pK_{P2}	pK_a constant for $H_2PO_4^-$ dissociation
pK_W	pK_a constant for water dissociation
pK_Z	pK_a constant for the unspecified compound dissociation
q_{Gal}	volumetric galactose secretion rate (C-mol L^{-1} h^{-1})
q_{NH}	volumetric ammonia consumption rate (mol L^{-1} h^{-1})
q_P	volumetric lactic acid secretion rate (C-mol L^{-1} h^{-1})

q_{Ph}	volumetric phosphoric acid consumption rate (mol L^{-1} h^{-1})
q_S	volumetric substrate consumption rate (C-mol L^{-1} h^{-1})
q_X	volumetric biomass growth rate (C-mol L^{-1} h^{-1})
T	temperature in the bioreactor (K)
t	time variable (h)
t_{lag}	lag-time coefficient (h)
Y_{gal}	galactose yield (g g^{-1})
z_i	charge number of the i-th ion
α	growth related production coefficient of lactic acid (g g^{-1})
μ_{max}	maximum specific growth rate (h^{-1})
σ	standard deviation
σ_{pH}	spread parameter is the gaussian pH function

References

1. Rantanen J, Khinast J. The Future of Pharmaceutical Manufacturing Sciences. *J Pharm Sci*. 2015;104(11):3612-3638. doi:10.1002/jps.24594.

2. Mears L, Stocks SM, Albaek MO, Sin G, Gernaey K V. Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends Biotechnol*. 2017;35(10):914-924. doi:10.1016/j.tibtech.2017.07.002.

3. Lara AR, Galindo E, Ramírez OT, Palomares L a. Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol Biotechnol*. 2006;34(3):355-381. doi:10.1385/MB:34:3:355.

4. Langheinrich C, Nienow AW. Control of pH in large-scale, free suspension animal cell bioreactors: Alkali addition and pH excursions. *Biotechnol Bioeng*. 1999;66(3):171-179. doi:10.1002/(SICI)1097-0290(1999)66:3<171::AID-BIT5>3.0.CO;2-T.

5. de Jonge LP, Buijs NAA, ten Pierick A, et al. Scale-down of penicillin production in Penicillium chrysogenum. *Biotechnol J*. 2011;6(8):944-958. doi:10.1002/biot.201000409.

6. Bylund F, Collet E, Enfors S-O, Larsson G. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Eng*. 1998;18(3):171. doi:10.1007/s004490050427.

7. Enfors SO, Jahic M, Rozkov A, et al. Physiological responses to mixing in large scale bioreactors. *J Biotechnol*. 2001;85(2):175-185. doi:10.1016/S0168-1656(00)00365-5.

8. Zou X, Xia J, Chu J, Zhuang Y, Zhang S. Real-time fluid dynamics investigation and physiological response for erythromycin fermentation scale-up from 50 L to 132 m3 fermenter. *Bioprocess Biosyst Eng*. 2012;35(5):789-800. doi:10.1007/s00449-011-0659-z.

9. Yang Y, Xia J, Li J, et al. A novel impeller configuration to improve fungal physiology performance and energy conservation for cephalosporin C production. *J Biotechnol*. 2012;161(3):250-256. doi:10.1016/j.jbiotec.2012.07.007.

10. Wang G, Tang W, Xia J, Chu J, Noorman H, van Gulik WM. Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses. *Eng Life Sci*. 2015;15(1):20-29. doi:10.1002/elsc.201400172.

11. Schmalzriedt S, Jenne M, Mauch K, Reuss M. Integration of Physiology and Fluid Dynamics. *Adv Biochem Eng Biotechnol*. 2003;80:19-68. doi:10.1007/3-540-36782-

9_2.

12. Haringa C, Tang W, Deshmukh AT, et al. Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines. *Eng Life Sci*. 2016;16(7):652-663. doi:10.1002/elsc.201600061.

13. Morchain J, Gabelle J-C, Cockx A. A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. *AIChE J*. 2014;60(1):27-40. doi:10.1002/aic.14238.

14. Haringa C, Tang W, Wang G, et al. Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chem Eng Sci*. 2018;175:12-24. doi:10.1016/j.ces.2017.09.020.

15. Le Moullec Y, Gentric C, Potier O, Leclerc JP. Comparison of systemic, compartmental and CFD modelling approaches: Application to the simulation of a biological reactor of wastewater treatment. *Chem Eng Sci*. 2010;65(1):343-350. doi:10.1016/j.ces.2009.06.035.

16. Vrábel P, Van der Lans RGJM, Van der Schot FN, Luyben KCAM, Xu B, Enfors SO. CMA: Integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations. *Chem Eng J*. 2001;84(3):463-474. doi:10.1016/S1385-8947(00)00271-0.

17. Vlaev D, Mann R, Lossev V, Vlaev SD, Zahradnik J, Seichter P. Macro-mixing and streptomyces fradiae modelling oxygen and nutrient segregation in an industrial bioreactor. *Chem Eng Res Des*. 2000;78(3):354-362. doi:10.1205/026387600527473.

18. Alvarado A, Vedantam S, Goethals P, Nopens I. A compartmental model to describe hydraulics in a full-scale waste stabilization pond. *Water Res*. 2012;46(2):521-530. doi:10.1016/j.watres.2011.11.038.

19. Rehman U, Audenaert W, Amerlinck Y, Maere T, Arnaldos M, Nopens I. How well-mixed is well mixed? Hydrodynamic-biokinetic model integration in an aerated tank of a full-scale water resource recovery facility. *Water Sci Technol*. 2017;76(8):1950-1965. doi:10.2166/wst.2017.330.

20. Hansen G, Johansen CL, Marten G, Wilmes J, Jespersen L, Arneborg N. Influence of extracellular pH on growth, viability, cell size, acidification activity, and intracellular pH of Lactococcus lactis in batch fermentations. *Appl Microbiol Biotechnol*. 2016;100(13):5965-5976. doi:10.1007/s00253-016-7454-3.

21. Cortés JT, Flores N, Bolívar F, Lara AR, Ramírez OT. Physiological effects of pH gradients on Escherichia coli during plasmid DNA production. *Biotechnol Bioeng*. 2016;113(3):598-611. doi:10.1002/bit.25817.

22. Paul EL, Atiemo-Obeng VA, Kresta SM, eds. *Handbook of Industrial Mixing*. Hoboken, NJ, USA: John Wiley \& Sons, Inc.; 2003. doi:10.1002/0471451452.

23. Spann R, Roca C, Kold D, Eliasson Lantz A, Gernaey K V., Sin G. A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. *Biochem Eng J*. 2018;135:49-60. doi:10.1016/j.bej.2018.03.016.

24. Villadsen J, Nielsen J, Lidén G. *Bioreaction Engineering Principles*. Boston, MA: Springer US; 2011. doi:10.1007/978-1-4419-9688-6.

25. Åkerberg C, Hofvendahl K, Hahn-Hägerdal B, Zacchi G. Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour. *Appl Microbiol Biotechnol*. 1998;49(6):682-690. doi:10.1007/s002530051232.

26. Aghababaie M, Khanahmadi M, Beheshti M. Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures. *Food Bioprod Process*. 2015;94(April):657-667. doi:10.1016/j.fbp.2014.09.007.

27. Schepers AW, Thibault J, Lacroix C. Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. *Enzyme Microb Technol*. 2002;30(2):187-194. doi:10.1016/S0141-0229(01)00466-5.

28. Peng RY, Yang TCK, Wang H, Lin Y, Cheng C. Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model. *J Chinese Agric Chem Soc*. 1997;35(5):485-494.

29. Musvoto EV, Wentzel MC, Loewenthal RE, Ekama GA. Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems. *Water Res*. 2000;34(6):1857-1867. doi:10.1016/S0043-1354(99)00334-6.

30. Loewenthal RE, Ekama GA, Marais GR. Mixed weak acid/base systems Part I - Mixture characterisation. *Water SA*. 1989;15(1):3-24.

31. Dawson RMC. *Data for Biochemical Research*. Oxford: Clarendon Press; 1969.

32. Davies CW. *Ion Association*. Londond: Butterworth; 1962.

33. McKay MD, Beckman RJ, Conover WJ. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*. 1979;21(2):239-245. doi:10.1080/00401706.1979.10489755.

34. Sin G, Gernaey K V., Neumann MB, van Loosdrecht MCM, Gujer W. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Res*. 2009;43(11):2894-2906. doi:10.1016/j.watres.2009.03.048.

35. Delafosse A, Collignon ML, Calvo S, et al. CFD-based compartment model for description of mixing in bioreactors. *Chem Eng Sci*. 2014;106:76-85. doi:10.1016/j.ces.2013.11.033.

36. Sin G, Gernaey K V, Lantz AE. Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis. *Biotechnol Prog*. 2009;25(4):1043-1053. doi:10.1002/btpr.166.

37. Omlin M, Reichert P. A comparison of techniques for the estimation of model prediction uncertainty. *Ecol Modell*. 1999;115(1):45-59. doi:10.1016/S0304-3800(98)00174-4.

Supplementary material

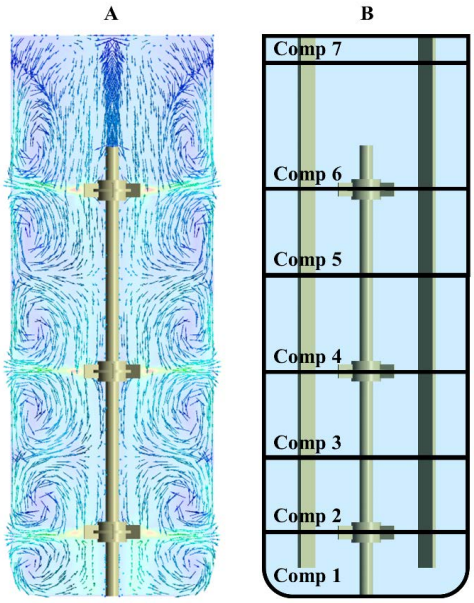


FIGURE S1. Design of the CFD- based compartment model for a stirred speed of 240 rpm. Velocity streamlines of the steady state solution (A). The seven compartments for the 700 L bioreactor (B).

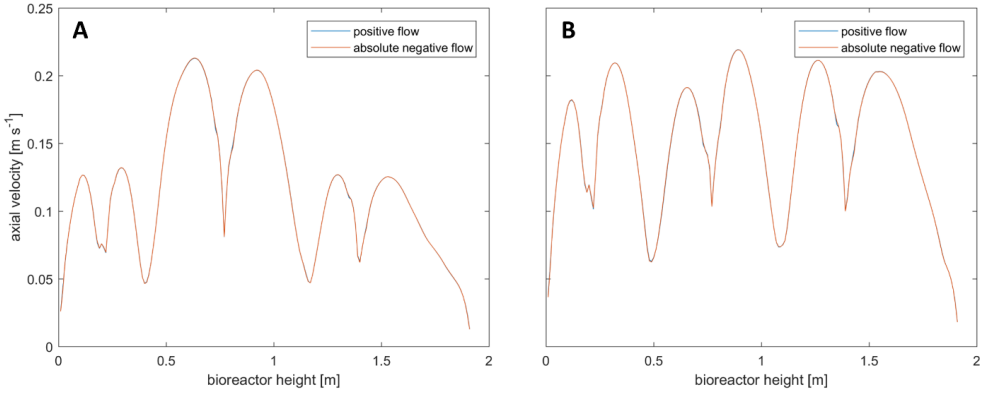


FIGURE S2. Axial velocities across the height of the bioreactor. For the steady state CFD simulation representing 130 rpm (A), and 240 rpm (B). Mean axial velocities were calculated over horizontal planes that were placed every 1 cm across the height of the bioreactor.

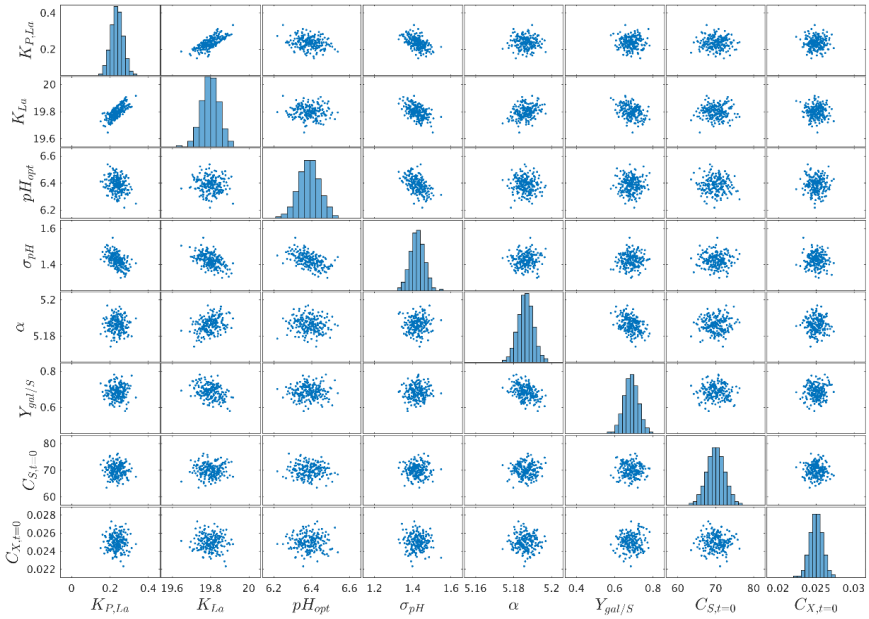


FIGURE S3. Sampling matrix for the input uncertainty space. The Latin Hypercube Sampling (LHS) technique and the Iman Conover rank correlation method were used to sample 200 independent inputs and to induce the known covariance matrix, respectively.

Paper G

On-line Process Risk Assessment of a 700 L Lactic Acid Bacteria Cultivation

Robert Spann, Krist V. Gernaey, Gürkan Sin

Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark

Submitted to Frontiers in Bioengineering and Biotechnology.

Abstract

An on-line process risk assessment framework was developed and applied for a 700 L *Streptococcus thermophilus* cultivation. To achieve this, a probabilistic soft sensor was applied that included a validated mechanistic biokinetic model and a CFD-based compartment model. The risk assessment framework considered imperfect mixing conditions (hence pH gradients) and its impact on the process performance. The process risk, defined as the likelihood of not achieving the target biomass production per batch, was calculated continuously during the cultivation process using the on-line measurements, namely ammonia solution addition and the pH measurement. A Monte Carlo simulation was performed each time the model was updated to account for uncertainties in the model parameters, namely the initial process conditions, and on-line measurements. The soft sensor predicted pH gradients ranging from pH 5.8 ± 0.02 at the top of the bioreactor to 6.1 ± 0.02 at the bottom of the bioreactor. In the present cultivation, the estimated process risk was to lose less than 140 g biomass per batch, which is ca. 3.5 % of the total production capacity. A sensitivity analysis indicated that the on-line ammonia addition had the highest impact on the risk quantification. This process analytical technology (PAT) tool could be included in an automated risk-based control framework that minimizes the risk of not achieving the economic objectives of the process.

1 Introduction

More and more scientific and risk-based methodologies have been implemented in pharmaceutical and related processes since the publications of the process analytical technology (PAT) guidance (FDA, 2004) and the quality by design (QbD) approach (ICH Q8(R2), 2009) (Rantanen & Khinast, 2015). These methodologies assist the industry to understand the manufacturing process and to control the process in a way that the quality of the product is assured by design. In the quality by design approach, the desired product attributes, such as purity, stability, and concentration, are defined, and critical quality attributes (CQAs) are identified. Critical process parameters (CPPs), i.e. process parameters that have an impact on the CQAs, are then determined based on process characterization studies (Rathore & Winkle, 2009). CPPs may include temperature, pH, feed flow rate etc. The acceptable range of the CPPs is defined as the design space that leads to the desired product quality. During production, these parameters need to be controlled by the PAT system and maintained within the design space to ensure a robust process operation and to ensure product quality in bioprocesses (Gnoth, Jenzsch, Simutis, & Lübbert, 2007).

Models are implemented to predict the CQAs by using the measured CPPs as model inputs in the framework of PAT (Glassey et al., 2011). Commonly, statistical models such as multivariate data analysis are applied to predict the effect of the CPPs on the CQAs (Mercier, Diepenbroek, Wijffels, & Streefland, 2014; Rathore, Mittal, Pathak, & Mahalingam, 2014). Nevertheless, mechanistic models and hybrid models (a combination of mechanistic and data-driven modelling techniques) are used as well (Kager, Herwig, & Stelzer, 2018; Mears, Stocks, Albaek, Sin, & Gernaey, 2017; Solle et al., 2017; Sommeregger et al., 2017). Since the CQAs can be hardly measured in real time, the models are especially beneficial in a soft sensor for on-line monitoring and control of industrial processes (Mandenius & Gustavsson, 2015). They enable to follow the dynamics of the CQAs in real time and to control the process accordingly.

Traditionally, risk assessment is conducted in the process design phase to identify process parameters with a high risk, which are then further investigated for process characterization (Rantanen & Khinast, 2015; Stocker, Toschkoff, Sacher, & Khinast, 2014). Risk management methods such as Failure Mode Effects Analysis (FMEA) provide a method to evaluate these risks (ICH Q9, 2005). The risk is thereby weighted based on the severity, occurrence and

detection of a process failure, i.e. a deviation from the parameter's nominal operating point. The severity is a measure for the seriousness of the consequences (with respect to the target product) if such a process failure happens; the occurrence is the expected probability of this event; and the detection indicates to which extent this process failure can or cannot be detected timely before the product is used (ICH Q9, 2005).

To date, risk is quantified statically during the process design phase (Adam, Suzzi, Radeke, & Khinast, 2011) but often not quantified dynamically in real time while a process is running. To achieve on-line risk quantification, we applied a Monte Carlo simulation in a model-based soft sensor including a CFD-based compartment model for on-line monitoring and risk quantification of a 700 L lactic acid bacteria cultivation. Lactic acid bacteria cultures are produced in large-scale bioreactors to be used subsequently in the dairy industry e.g. for yogurt or cheese production. The compartment model was applied to account for heterogeneous process conditions (especially pH) in the process. In the Monte Carlo simulation, several uncertainties were considered: model parameter uncertainties, process input variations, and on-line measurement errors. The soft sensor predicted therefore a probability distribution of the state variables on-line, including the critical quality attribute for the case study defined as the biomass yield. The probability of not achieving the target biomass production and the corresponding risk were quantified based on the predicted probability distribution and updated on-line. A sensitivity analysis was then applied to provide better insight into which process parameters have been contributing to the computed process risk.

2 Materials and Methods

2.1 Cultivation conditions

A batch cultivation of the homolactic *S. thermophilus* (provided by Chr. Hansen A/S, Hørsholm, Denmark) was carried out in a stirred tank bioreactor at a stirring speed of 132 rpm, 40 °C, and with N_2 headspace gassing. The stirred tank bioreactor (Chemap AG, Switzerland) was equipped with three 6-blade Rushton turbines, had four baffles, and was filled with approx. 700 L cultivation medium initially (Figure 1 A). The pH was controlled by adding ammonia solution (24 % (w/v) NH_4OH) at the bottom of the bioreactor with the pH set point at 6. The pH controlling sensor was located 0.3 m above the bottom of the bioreactor close to the bioreactor wall. In addition, pH sensors (CPS471D, Endress+Hauser AG, Switzerland) measured the pH at heights of 0.1 m and 1.6 m with a distance of 0.1 m to the reactor wall. The initial pH was 6.8. The cultivation medium contained 70 g L^{-1} lactose, 10 g L^{-1} casein hydrolysate, 12 g L^{-1} yeast extract, 11.5 mM K_2HPO_4, 36.6 mM sodium acetate, 8.2 mM trisodium citrate, 0.8 mM $MgSO_4$, and 0.3 mM $MnSO_4$.

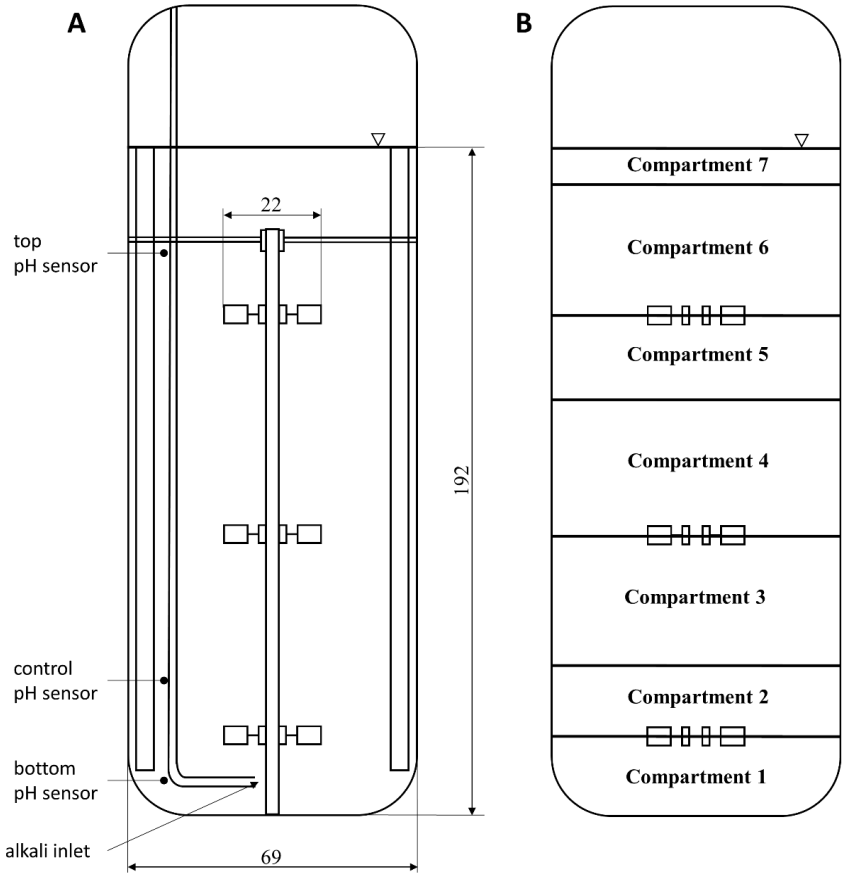


Figure 1. Simplified bioreactor setup (A) and the compartment model (B). A bioreactor with three six-blade Rushton turbines, four baffles, and a liquid volume of 700 L was used. The base (ammonia solution) was added at the bottom of the bioreactor. B: 7 compartments were designed based on the axial velocities of the steady state CFD solution resembling a stirrer speed of 132 rpm (for details see Spann et al., 2018, submitted).

2.2 Design and validation of the compartment model

The compartment model has been designed and validated elsewhere (manuscript in preparation). The compartment model was based on a computational fluid dynamics simulation that had been validated with tracer pulse experiments earlier (Spann et al., 2018, submitted). In accordance to the six recirculation loops that are built by the three Rushton turbines (Vrabel, Van Der Lans, Luyben, Boon, & Nienow, 2000), six compartments were defined, where each compartment resembled a recirculation loop (Figure 1 B). In addition, a seventh compartment was defined at the top zone of the bioreactor, where low axial velocities were predicted. The compartments were defined based on the axial velocity, i.e. the compartments were separated at the positions of the lowest axial velocities. This methodology assumed fast circumferential mixing, which might be true for the inner radius of the bioreactor, as the impellers turn with 2

rounds per second. However, it simplified the fluid dynamics close to the bioreactor wall and baffles, where slow circumferential mixing was

Table I. Properties of the compartment model (half of the bioreactor was modelled).

Compartment interconnection	Interface area [m^2]	Velocity [m s^{-1}]	Compartment no.	Volume [m^3]
1 ↔ 2	0.1754	0.0693	1	0.0388
2 ↔ 3	0.1839	0.0476	2	0.0364
3 ↔ 4	0.1754	0.0810	3	0.0671
4 ↔ 5	0.1839	0.0527	4	0.0768
5 ↔ 6	0.1754	0.0669	5	0.0396
6 ↔ 7	0.1847	0.0541	6	0.0806
			7	0.0191

predicted by the CFD simulation (Spann et al., 2018, submitted). The pros and cons of this compartment design are discussed elsewhere (manuscript in preparation). However, the focus of this article are the applications of a compartment model with respect to on-line risk assessment, and not the detailed discussion of the compartment model. The volumes of the compartments and the flows between the compartments (Table I) were extracted from the CFD model that was built in ANSYS CFX 17.1 (ANSYS, Inc., US-PA). Half of the bioreactor was modelled as there was a vertical symmetry in the middle of the bioreactor. The compartment model was implemented together with the kinetic model (see below) in MATLAB® R2017a (The MathWorks®, Natick, MA).

2.3 Biokinetic model and pH simulation

The dynamic model comprised a biological and a chemical model part as described in detail in Spann et al. (2018). The biokinetic model predicted the evolution of the state variables, such as biomass, lactic acid, and lactose concentration. The chemical model was a mixed weak acid/base model describing the dissociation reactions of the charged components, such as ammonium and lactate.

2.3.1 The biokinetic model

The biokinetic model was based on the global stoichiometric process equation (Villadsen, Nielsen, & Lidén, 2011) (Eq. (1)-(2)). The biomass growth rate was modelled as a function that depended on the lag-time (f_{lag}), lactose inhibition and limitation (f_S) (Åkerberg, Hofvendahl, Hahn-Hägerdal, & Zacchi, 1998), lactate inhibition(f_P) (Aghababaie, Khanahmadi, & Beheshti, 2015), and the pH in the cultivation broth (f_{pH}) (Schepers, Thibault, & Lacroix, 2002) (Eq.(3)-(4)).

$$\boldsymbol{Lactose + Ammonia + Phosphoric\ acid \rightarrow Biomass + Lactic\ acid + Galactose} \quad (1)$$

$$\mathbf{q_S \cdot CH_2O + q_{NH} \cdot NH_3 + q_{Ph} \cdot H_3PO_4 \rightarrow q_X \cdot CH_aO_bN_cP_d + q_P \cdot CH_2O + q_{Gal} \cdot CH_2O} \quad (2)$$

$$\frac{\boldsymbol{dC_X}}{\boldsymbol{dt}} = \mu_{\boldsymbol{max}} \cdot \boldsymbol{f_{lag} \cdot f_S \cdot f_P \cdot f_{pH} \cdot C_X} \quad (3)$$

$$\frac{\boldsymbol{dC_X}}{\boldsymbol{dt}} = \mu_{\boldsymbol{max}} \cdot \left(\mathbf{1} - \boldsymbol{e}^{-\frac{\boldsymbol{t}}{\boldsymbol{t_{lag}}}}\right) \cdot \frac{\boldsymbol{C_S}}{\boldsymbol{C_S + K_s + \frac{C_S^2}{K_I}}} \cdot \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{\boldsymbol{K_{P,La}(C_{LA} - K_{La1})}}} \cdot \boldsymbol{e}^{-\left(\frac{(\boldsymbol{pH_{opt} - pH})^2}{\boldsymbol{\sigma_{pH}^2}}\right)} \cdot \boldsymbol{C_X} \quad (4)$$

with K_{La1} dependent on the pH:

$$\boldsymbol{K_{La1} = K_{La} \cdot \frac{1}{1 + e^{K_{P,pH1} * (pH - K_{P,pH2})}}} \quad (5)$$

The lactic acid synthesis was considered to be growth dependent (Peng, Yang, Wang, Lin, & Cheng, 1997):

$$\boldsymbol{\frac{dC_P}{dt} = \alpha \cdot \frac{dC_X}{dt}} \quad (6)$$

The lactose consumption rate was the sum of the biomass growth and the lactic acid synthesis rate considering the secretion of galactose (Y_{gal}) since the used strain metabolizes only glucose and secretes galactose under the present cultivation conditions:

$$\boldsymbol{\frac{dC_S}{dt} = -\left(1 + Y_{gal}\right) \cdot \left(\frac{dC_X}{dt} + \frac{dC_P}{dt}\right)} \quad (7)$$

The kinetic parameters were estimated from the data obtained in five lab-scale cultivations, and validated with an independent data set. The experiments were conducted under different substrate (20 and 70 g L^{-1}) and pH conditions (5.5 <= pH <= 7.0) including identifiability and uncertainty analysis (Spann et al., 2018). The derived parameters including the uncertainty of the estimated parameter values are listed in Table II.

Table II Kinetic parameters of the dynamic model for the *S. thermophilus* cultivation.

Symbol	Value	Std. deviation	Reference
Biological model			
K_I	164 g L^{-1}	n.d.	(Åkerberg et al., 1998)
K_{La}	19.80 g L^{-1}	0.05 g L^{-1}	(Spann et al., 2018)
$K_{P,La}$	0.24 L g^{-1}	0.03 L g^{-1}	(Spann et al., 2018)
$K_{P,pH1}$	20	n.d.	(Spann et al., 2018)
$K_{P,pH2}$	7	n.d.	(Spann et al., 2018)
K_S	0.79 g L^{-1}	n.d.	(Åkerberg et al., 1998)
pH_{opt}	6.39	0.06	(Spann et al., 2018)
t_{lag}	updated in the soft sensor		
Y_{gal}	0.69 g g^{-1}	0.04 g g^{-1}	(Spann et al., 2018)
α	5.19 g g^{-1}	0.01 g g^{-1}	(Spann et al., 2018)
μ_{max}	Initial value: 2.06 h^{-1}, updated in the soft sensor		(Spann et al., 2018)
σ_{pH}	1.42	0.04	(Spann et al., 2018)
Mixed weak acid/base model			
$K'_{r,C1}$	10^7 s^{-1}		(Musvoto et al., 2000)
$K'_{r,LA}$	10^7 s^{-1}		(Musvoto et al., 2000)
$K'_{r,NH}$	10^{12} s^{-1}		(Musvoto et al., 2000)
$K'_{r,P1}$	10^8 s^{-1}		(Musvoto et al., 2000)
$K'_{r,P2}$	10^{12} s^{-1}		(Musvoto et al., 2000)
$K'_{r,W}$	10^{10} s^{-1}		(Musvoto et al., 2000)
$K'_{r,Z}$	10^7 s^{-1}		(Musvoto et al., 2000)
pK_{C1}	$3404.7/(T - 14.8435 + 0.03279 \cdot T)$		(Loewenthal et al., 1989)
pK_{LA}	3.86		(Dawson, 1969)
pK_{NH}	$2835.8/(T - 0.6322 + 0.00123 \cdot T)$		(Loewenthal et al., 1989)
pK_{P1}	$799.3/(T - 4.5535 + 0.01349 \cdot T)$		(Loewenthal et al., 1989)
pK_{P2}	$1979.5/(T - 5.3541 + 0.01984 \cdot T)$		(Loewenthal et al., 1989)
pK_W	14		(Loewenthal et al., 1989)
pK_Z	9.4		(Spann et al., 2018)
T	313 K		measured process condition
Initial Conditions			
$C_{Gal,t=0}$	0.0 g L^{-1}		
$C_{Glc,t=0}$	0.0 g L^{-1}		
$C_{P,t=0}$	0.0 g L^{-1}		
$C_{S,t=0}$	70 g L^{-1}	2.3 g L^{-1}	
$C_{tCO,t=0}$	$1.002 \cdot 10^{-5}$ $mol\ L^{-1}$		
$C_{tNH,t=0}$	0.005 g L^{-1}		
$C_{tPh,t=0}$	2 g L^{-1}		
$C_{tZ,t=0}$	2 mol L^{-1}		
$C_{X,t=0}$	0.025 g L^{-1}	$8 \cdot 10^{-4}$ g L^{-1}	

2.3.2 The mixed weak acid/base model

The objective of the mixed weak acid/base model was to predict the pH (as the negative logarithm of the hydrogen ion activity: $pH = -\log_{10}\{H^+\}$). To this end, this model part comprised the dissociation reactions of the charged components in the cultivation (Musvoto, Wentzel, Loewenthal, & Ekama, 2000), such as ammonium, lactate, phosphate, carbonate, etc. which are relevant in the investigated pH range (Table III) (Spann et al., 2018). The pKa values were derived from Dawson (1969) and (Loewenthal, Ekama, & Marais, 1989) (Table II), and the activity coefficients were calculated by a modified Debye-Hückel model by Davies (Davies, 1962):

$$log(f_i) = -1.825 \cdot 10^6 \cdot (78.3 \cdot T)^{-1.5} \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - 0.3 \cdot I \right) \quad (8)$$

With the ionic strength (I):

$$I = \frac{1}{2} \sum_i z_i^2 C_i \quad (9)$$

A P-controller with a controller gain (K_P) of (5 mol L^{-1} · liquid volume [L]) was applied to maintain the pH at the set point value of 6 by adding ammonia solution:

$$NH_4OH_{add} = K_P \cdot (pH_{set} - pH) \quad (10)$$

The model was implemented and solved in MATLAB. The numerical solver ode15s was used because the present model contains slow (e.g. the biomass growth rate) and fast time constants (e.g. the ammonia dissociation rate and the flow rates between the compartments) resulting in a stiff system of ordinary differential equations. The implemented stoichiometric matrix of the biological and chemical model can be found in the Supplementary Material.

Table III. Kinetics of the mixed weak acid/base model. f_m and f_d are the mono- and divalent activity coefficients, respectively; see Loewenthal et al. (1989) and Musvoto et al. (2000).

Dissociation process	Reaction	reaction rate vector	apparent equilibrium constant
Ammonium	$NH_4^+ \leftrightarrow NH_3 + H^+$	$K'_{r,NH} \cdot K'_{NH} \cdot [NH_4^+] - K'_{r,NH} \cdot [NH_3] \cdot [H^+]$	$K'_{NH} = 10^{-pK_{NH}}$
Phosphate 1	$H_3PO_4 \leftrightarrow H_2PO_4^- + H^+$	$K'_{r,P1} \cdot K'_{P1} \cdot [H_3PO_4] - K'_{r,P1} \cdot [H_2PO_4^-] \cdot [H^+]$	$K'_{P1} = 10^{-pK_{P1}}/f_m^2$
Phosphate 2	$H_2PO_4^- \leftrightarrow HPO_4^{2-} + H^+$	$K'_{r,P2} \cdot K'_{P2} \cdot [H_2PO_4^-] - K'_{r,P2} \cdot [HPO_4^{2-}] \cdot [H^+]$	$K'_{P2} = 10^{-pK_{P2}}/f_d$
Carbonate 1	$H_2CO_3^* \leftrightarrow HCO_3^- + H^+$	$K'_{r,C1} \cdot K'_{C1} \cdot [H_2CO_3^*] - K'_{r,C1} \cdot [HCO_3^-] \cdot [H^+]$	$K'_{C1} = 10^{-pK_{C1}}/f_m^2$
Lactate	$C_3H_6O_3 \leftrightarrow C_3H_5O_3^- + H^+$	$K'_{r,LA} \cdot K'_{LA} \cdot [C_3H_6O_3] - K'_{r,LA} \cdot [C_3H_5O_3^-] \cdot [H^+]$	$K'_{LA} = 10^{-pK_{LA}}/f_m^2$
Water	$H_2O \leftrightarrow OH^- + H^+$	$K'_{r,W} \cdot K'_W - K'_{r,W} \cdot [OH^-] \cdot [H^+]$	$K'_W = 10^{-pK_W}/f_m^2$
Unknown compound	$ZH^+ \leftrightarrow Z + H^+$	$K'_{r,Z} \cdot K'_Z \cdot [ZH^+] - K'_{r,Z} \cdot [Z] \cdot [H^+]$	$K'_Z = 10^{-pK_Z}/f_m^2$

2.4 Probabilistic soft sensor for on-line monitoring

The aim of the probabilistic soft sensor is to predict the measurable and unmeasurable process variables, such as the biomass and substrate concentration, and the pH in real time. The algorithm for the probabilistic sensor is shown in Table IV and the details of the soft sensor including a validation with 2 L lab-scale experiments can be found elsewhere (Spann et al., 2018).

Once the process is started, the soft sensor is updated in 5 min intervals (Table IV). The initial process conditions are defined as specified for the real cultivation (Table IV, step 1-2). The soft

sensor uses the latest on-line measurements of the process, namely the added ammonia quantity and the pH (Table IV, step 3) to update the model parameters μ_{max} and t_{lag} (Table IV, step 4). The parameters are updated in 5 min intervals, and are then used as input to the dynamic model that predicts both the current value and the future course of the state variables (Table IV, step 5). In this study, the soft sensor was applied off-line once the cultivation was performed for demonstration purposes. The on-line measurements were hereby used as they would be available on-line. The off-line measurements were only used to assess the goodness of the model fit (see below) but not to update the soft sensor.

Table IV. Methodology of the probabilistic soft sensor

Step 1	**Define the initial process conditions (x_0) of the real process and model**	
Step 2	**Start the cultivation**	
Iterate step 3 to 7 in 5 minutes intervals until t_{end} (cultivation completion)		
Step 3	**Read on-line measurements**	
	pH and ammonia addition rate (pH, $q_{NH,add}$)	
Step 4	**Update the kinetic parameters μ_{max} and t_{lag}**	
	Step 4.1	Data reconciliation
		$NH_4^+ + C_3H_5O_3^- = q_{NH,add} + q_P = 0$
		$q_X = q_X / a$
	Step 4.2	Parameter update
		$\mu_{max,k} = \frac{q_{X,updated}}{f_{lag,k-1} \cdot f_{S,k-1} \cdot f_{P,k-1} \cdot f_{pH,k-1} \cdot X_{k-1}}$
		$t_{lag,k} = t_{lag,k-1} + (t_{pH=6,measured} - t_{pH=6,predicted})$
Step 5	**Monte Carlo simulation of the model**	
	Step 5.1	Define the input uncertainty space (once/ not every interval) (σ_θ,σ_{x_0})
	Step 5.2	Sample the independent input matrix (once/ not every interval)
		SAMPLE MATRIX ($\Theta_{l \times N}$)
	Step 5.3	Monte Carlo simulation
		for 1:N
		Solve y(t) = Model(θ_j,x_0)
Step 6	**Process risk quantification**	
	$process\ risk = \sum_m consequence \cdot likelihood$	
Step 7	**Save current state**	

A Monte Carlo simulation of the dynamic model is performed every interval as explained in detail in Spann et al. (2018) (Table IV, step 5). To this end, the input uncertainties are first identified and defined. Second, random input samples are generated, and third, the Monte Carlo simulation is performed. In this study, uncertainties in the model parameters, initial conditions,

and the ammonia addition are considered. The Latin hypercube sampling technique is used to generate N = 200 random samples (Supplementary Figure S1) from the input uncertainty domain (M. D. McKay, Beckman, & Conover, 1979; Sin, Gernaey, Neumann, van Loosdrecht, & Gujer, 2009). 200 model simulations were therefore performed every interval that the soft sensor was updated providing a probability distribution of the model outputs. The model predictions of the biomass production were then assessed for the on-line risk quantification.

2.5 Assessment of the soft sensor predictions

The quality of the soft sensor predictions was assessed with the root mean sum of squared errors (RMSSE) with respect to the off-line measurements:

$$\boldsymbol{RMSSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i}}^{\boldsymbol{n}} \left(\boldsymbol{y}_{\boldsymbol{meas,i}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}}\right)^{\mathbf{2}}} \tag{11}$$

2.6 Process risk quantification

The risk of not achieving the target production of biomass was calculated on-line as a result of the soft sensor predictions (Table IV, step 6). The biomass was selected because the lactic acid bacteria were the desired product of this process. The target biomass production was defined as 4410 g biomass per batch that was based on previous 2 L lab-scale experiments (see the Supplementary Material for the detailed calculation). The loss/surplus (here named consequence) for each of the j Monte Carlo simulation predictions was then calculated as the difference between the model prediction ($\hat{y}$) and the target:

$$\boldsymbol{consequence}_{\boldsymbol{j}} = \hat{\boldsymbol{y}}_{\boldsymbol{j}} - \boldsymbol{target} \tag{12}$$

Risk is generally defined in the process industries as the likelihood of an undesirable event (u.e.) times the consequence of that event (Cameron & Raman, 2005). The risk of several undesirable events is consequently the sum of their individual risks (Eq. (13)). In this study, the consequence of an undesired event was the loss of the biomass in terms of total biomass amount per batch (Eq. (12)). The likelihood of this event was the probability of this event that was predicted by the Monte Carlo simulation.

$$\boldsymbol{process\ risk} = \sum_{\boldsymbol{m}} \boldsymbol{consequence}(\boldsymbol{u.e.}) \cdot \boldsymbol{likelihood}(\boldsymbol{u.e.}) \tag{13}$$

where m ϵ j is the number of undesirable events (u.e.).

2.7 Sensitivity analysis

A linear regression on the Monte Carlo simulation outputs (see above) was performed to measure the sensitivity of the parameters on the model outputs (Saltelli et al., 2008). This method is also called Standardized Regression Coefficients (SRC) method. In this study, the final biomass concentration after 6 h of cultivation was the model output of interest, as the biomass was the target product of the cultivation. A first order linear multivariate model was therefore fitted to the j Monte Carlo simulation predictions of the final biomass concentration using the constrained linear least squares algorithm (lsqlin) in MATLAB:

$$\hat{y}_{X,j} = a + \sum_i b_i \cdot \hat{\theta}_{i,j} \tag{14}$$

where $\hat{y}_X$ represents the predicted biomass concentration (in this case study, it was chosen the prediction after 6 h of cultivation which is the typical length of batch cultivation), a the ordinate-intercept of the linear model, b the linear regression coefficient, $\hat{\theta}$ the parameter value, i the index of the model input of interest, and j the index of the Monte Carlo simulation.

The standardized regression coefficient (SRC_i) (also called β_i) was obtained by scaling the linear regression coefficient b_i using the standard deviation of the model inputs and output of the Monte Carlo simulation:

$$SRC_i = \frac{\sigma_{\hat{\theta}_i}}{\sigma_{\hat{y}_X}} \cdot b_i \tag{15}$$

The SRC_i^2 are then the relative variance contributions to the model output variance of the linear model. SRC_i can take values between -1 and 1, whereas a high absolute value indicates a large effect of the parameter i on the model output, a negative sign means a negative effect, and a positive sign indicates a positive effect on the output (Sin, Gernaey, Neumann, van Loosdrecht, & Gujer, 2011). If the coefficient of determination is sufficiently high, e.g. $R^2 > 0.7$, which implies that the model is sufficiently linear, the SRC_i is considered to be a valid measure for the sensitivity (Campolongo, 1997; Saltelli, Ratto, Tarantola, & Campolongo, 2006).

3 Results and Discussion

A validated model-based soft sensor was applied to predict unmeasurable attributes such as the biomass concentration in a lactic acid bacteria cultivation and to quantify the risk of not achieving the target biomass production. To this end, a CFD-based compartment model was used to provide a reliable risk quantification since there exist pH gradients in the 700 L bioreactor that occur due to insufficient mixing.

3.1 On-line pH gradient monitoring

The probabilistic soft sensor was applied to a historical cultivation data set, whereas the historical on-line data (pH and balance readout of the ammonia addition) were used as they would be available on-line. The soft sensor used the on-line data to update the model parameters μ_{max} and t_{lag} in 5 min intervals, as described in the Materials and Methods section and in Spann et al. (2018). A Monte Carlo simulation of the dynamic model was performed within the soft sensor to account for uncertainties in the model parameters, the ammonia addition quantity measurement, and the initial biomass inoculation and lactose concentration. The Monte Carlo simulation with 200 input samples accounting for uncertainties in the model parameters, initial process conditions, and the ammonia addition propagated the error to the model outputs, such as the pH (Figure 2 left column) and the biomass and lactose concentration (Figure 2 right column). The output of the soft sensor was therefore a probability distribution of the state variables, and the 95 % confidence intervals of the model predictions are shown (Figure 2). The predictions of the earlier, current, and future states of the system are shown as an example at different times: 2 h, 4 h, and 6 h (Figure 2 rows). The virtual implementation with 5 min intervals may be found in the Supplementary Movie.

The initial pH was ca. 6.8 and then dropped due to lactic acid secretion until the controlling pH value 6 was reached when the base addition started (Figure 2 left column). In the first 2.5 h, no ammonia was added since the pH > 6 at the controlling position (for the position of the pH controlling sensor see Figure 1). There was consequently also no pH gradient predicted until the ammonia solution was added, as the cell density was maintained homogeneously distributed in the liquid phase.

As soon as the base addition started, a pH gradient was formed (Figure 2 B and C): In the bottom compartment of the bioreactor where the ammonia solution was added, a pH of up to 6.1 ± 0.02 was predicted due to the base addition. In the compartment 6 (second compartment from the top of the bioreactor), the pH dropped down to 5.8 ± 0.02 during the exponential growth phase. The quality of the quantitative pH gradient prediction was not the scope of this paper and was assessed elsewhere (manuscript in preparation). However, the on-line predicted gradients seemed to cover the reality qualitatively, as pH gradients in the range of 5.9 to 6.3 have also been measured in this cultivation (Spann et al., 2018, submitted).

The results of the on-line prediction of the pH gradient could be used at the production scale to minimize the risk of faulty batches for example by (i) monitoring the extent of the gradients; (ii) controlling the process; and (iii) rethinking about an improved bioreactor, impeller, or base addition design. In case the soft sensor is implemented as a monitoring tool – as shown in this study – plant operators could manually supervise the process and take actions in case the pH gradients reach a critical level. They could take risk-based decisions as they have a measure for whether the mixing is sufficient with respect to the pH. The soft sensor could also be applied for automated on-line control. In order to avoid too large pH gradients, the impeller speed could be increased, for example. Apart from this, the cultivation temperature could also be altered, in order to regulate the biomass growth rate (which is not included in the presented model). A decreased biomass growth rate would indeed decrease the lactic acid production, and hence the pH gradient might decrease. However, this might also result in a longer cultivation time. A model-based control algorithm could be implemented to predict the best control strategy (Jiménez-Hornero, Santos-Dueñas, & García-García, 2009; Mears, Stocks, Sin, & Gernaey, 2017). In case severe gradients occur frequently, results of such a model might also be an incentive for the production department to re-evaluate the bioreactor design, especially parts like the impeller or base addition inlet that could be modified more easily than the bioreactor itself.

The probabilistic soft sensor predicted in addition to the pH gradients the biological state variables, such as the biomass, lactose (substrate), and lactic acid (data not shown) concentration (Figure 2 right column). With the cultivation time, the update of the parameters μ_{max} and t_{lag} improved the prediction. After 3 h, the RMSSE for the biomass concentration prediction – the target product – was smaller than 0.4 ± 0.1 g L^{-1} that corresponds to an error of less than 10 % with respect to the final biomass concentration (data not shown).

This soft sensor, if implemented at a production site, provides the plant operators with a PAT tool to monitor the course of the cultivation with biological variables instead of the base addition profiles that have little direct meaning. A further strength of the soft sensor is that it could be applied to predict the end time of the cultivation, i.e. when the target cell mass will be achieved (Petrides & Siletti, 2004). Downstream capacities, including primed machines and workforce, could be scheduled and prepared accordingly. Furthermore, also subsequent steps in the upstream process, such as cleaning or pre-cultures for subsequent cultivations could be optimally planned. Overall, it might reduce the downtime of the plant equipment leading to a more economical operation.

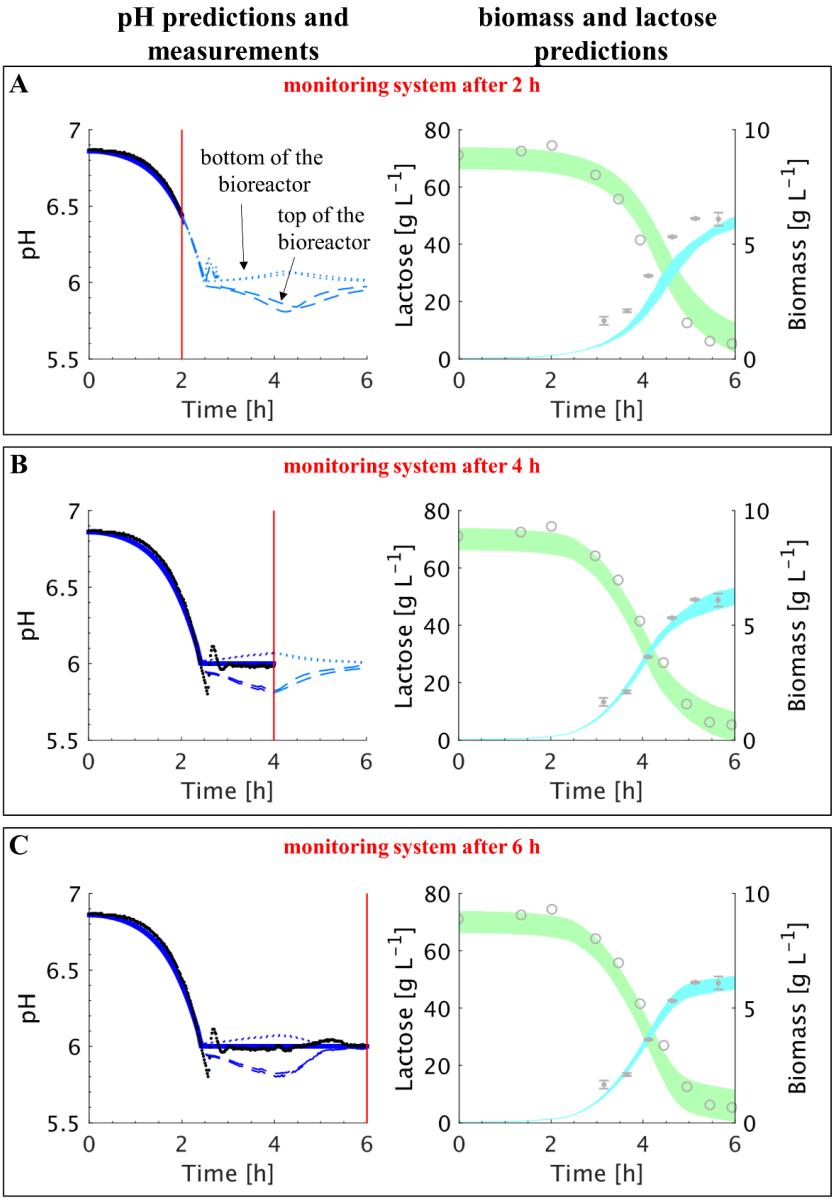


Figure 2. Probabilistic soft sensor to predict the pH gradient (left column), biomass growth, and substrate consumption (right column). The soft sensor using the compartment model was applied to data of a 700 L *S. thermophilus* batch cultivation. Predictions of the pH at the controlling position (blue line), the pH at the bottom compartment (blue dots), the pH in the top compartment (blue dashed line), biomass (cyan), and lactose (green) are shown. The off-line measurements for biomass (gray dots with standard deviation) and lactose (gray circles) are shown for comparison only, but were not used for the on-line update of the parameters.

3.2 On-line process risk quantification

A frequently asked question during the production process is, “What is the risk of not achieving the target yield (titer, productivity, etc.)?” In order to demonstrate the capability of the probabilistic soft sensor to quantify and update this risk while the process is running, we selected the biomass yield and total biomass production per batch as an example.

The target yield was defined to be 0.09 ± 0.003 g biomass per g lactose based on previous 2 L lab-scale experiments (data not shown). The target was to achieve at least the same yield when the process was scaled up to the 700 L bioreactor. The undesired event was therefore to achieve less than the target yield. The risk was considered as the loss of product (biomass) per batch. It is quantified as the sum of the likelihood of the undesirable events times the amount of lost product (see Materials and Methods).

In our case, we got the likelihood from the output of the Monte Carlo simulation that considered uncertainties in the model parameters, initial process conditions, and the ammonia solution addition balance readout. The output of the probabilistic soft sensor were 200 model predictions. We considered the probability distribution of the biomass concentration prediction. The risk quantification method will be first presented with the biomass concentration of the final model prediction after 6 h of the cultivation as an example. Subsequently, the results of the on-line risk quantification considering the dynamic model updates will be shown.

First, the biomass yield was calculated and a histogram of the predicted probability distribution is shown (Figure 3 A). The biomass yield distribution ranged from 0.076 to 0.096 g biomass (g lactose)$^{-1}$. Some simulations did not reach the target yield. Second, the cumulative distribution function of the predicted yield minus the target yield was calculated (Figure 3 B). In this example, the probability of not achieving the target yield was 75 %. In other words, 75 % of the 200 simulations predicted that the final biomass yield was smaller than the desired target yield. 25 % of the Monte Carlo predictions were accordingly equal or larger than the target yield. Third, the biomass production of the entire batch was calculated considering the bioreactor volume (Figure 3 C). The total production amount might be of higher interest for a company than the yield as the obtained mass is crucial for sale. Product quality aspects were not considered in this work but could of course be included in the model. The risk is equivalent with the area under the cumulative distribution function that corresponds to the undesired events (Figure 3 C, grey shaded area). In this example, the risk was the loss of 140 g biomass per batch.

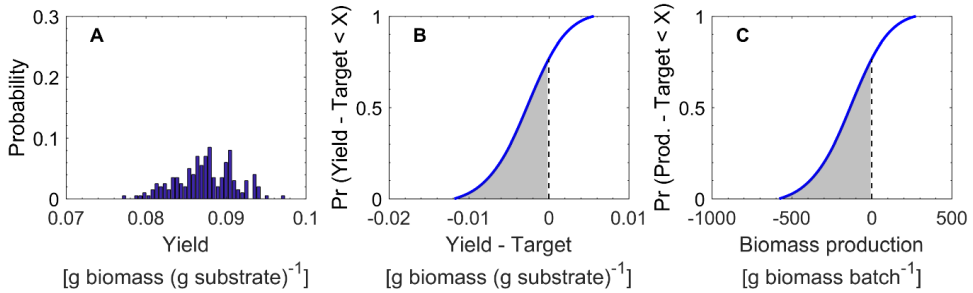


Figure 3. Probability distribution of the target biomass yield and production quantity. The probability distribution of the biomass yield after 6 h of cultivation as predicted by the Monte Carlo simulation (A); Cumulative distribution function of the yield with respect to the target yield (B); Cumulative distribution function of the total biomass production per batch (C). The grey shaded area under the cumulative distribution function represents the risk.

As a result of the probabilistic soft sensor, the risk could be predicted on-line considering the model updates (Figure 4). In this study, the predicted biomass concentration at the end of the cultivation (after 6 h) was considered. However, the time point and the desired product or other process attributes could be amended for other studies.

In the beginning of the cultivation, the risk could not be properly predicted as no or only little information from the on-line measurements was available (Figure 4, Initialization phase). As more on-line data was provided, the soft sensor could be updated, and hence the model predictions became more accurate. The on-line risk calculation needed therefore an initialization phase, waiting for enough on-line data (ammonia addition and pH measurements) to update the lag-time parameter and the maximum specific growth rate parameter. Once enough on-line data was available, a proper on-line risk quantification was achieved during the process operation. In this study, the initialization phase was set until t_{lag} was finally updated after 2 h and 25 min, when the base addition started. Nevertheless, the boundaries for the initialization phase need to be adapted in case the system would be applied for a different cultivation system.

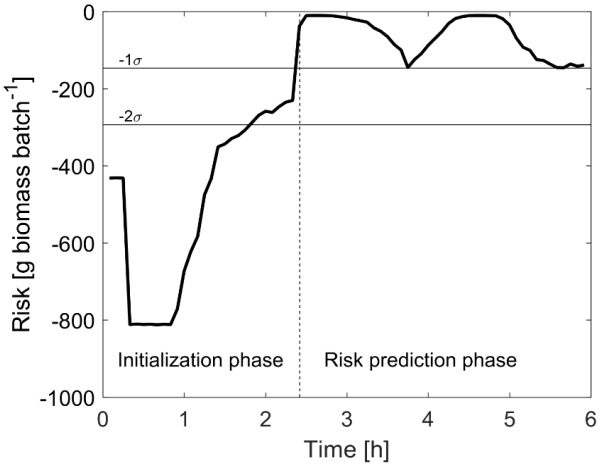


Figure 4. On-line risk quantification during the cultivation. The risk as biomass production loss per batch was quantified on-line based on the output of the probabilistic soft sensor that was updated in 5 min intervals. Limited on-line measurements were available in the beginning of the cultivation that did not enable a proper risk quantification (Initialization phase). With more on-line data, the dynamic model parameters could be updated in the soft sensor allowing the risk prediction (Risk prediction phase).

The risk was low, i.e. close to zero, when all Monte Carlo simulations achieved the target. In the present case, the risk is low after 2.5 h (Figure 4, Risk prediction phase). Later on, the risk became higher between 3-4 h reaching a predicted risk of 140 g biomass that could be lost per batch. Next, the risk was predicted to be low again after 4 h and 30 min, and finally, when the soft sensor was updated after 6 h cultivation, the risk was 140 g biomass per batch. This oscillating risk prediction could be attributed to the on-line update of the μ_{max} parameter that was dependent on the base addition. The base addition reflects the lactic acid secretion by the lactic acid bacteria and hence the biomass growth can be predicted. The on-line risk quantification captured therefore effects of the biomass growth rate. The growth rate changed in the presented cultivation. This resulted in the oscillations of the predicted biomass yield that

led to the oscillating risk prediction (Supplementary Figure S2). The predicted risk is nevertheless within the 1σ range of the target yield (0.09 ± 0.003 g biomass (g lactose)$^{-1}$) and regarded as natural variability of the process. This indicated that there was no yield decrease in the presented 700 L cultivation compared to the 2 L lab-scale experiments. However, replicates of the 700 L cultivations would be needed to validate the results statistically.

For an industrial application, the risk could be calculated as profit loss, i.e. an economic risk assessment (Gargalo, Cheali, Posada, Gernaey, & Sin, 2016; Hasanly, Khajeh Talkhoncheh, & Karimi Alavijeh, 2017), as the economic aspect might be the driving force for the production. The risk could then be quantified in e.g. $ per batch. Furthermore, possible loss of product quantity during the downstream operations could also be considered. The benefit from the monitoring system is that one can reflect and take action either by automated on-line control or manually, i.e. the action by a process operator. The operators could obtain an on-line measure to assess the risk of faulty batches and react accordingly, e.g. by increasing the stirrer speed to decrease pH gradients. In future, the soft sensor could be applied for on-line control, and hence controlling process parameters in such a way that the risk of losing product or profit remains as low as possible. To this end, it might also be necessary to include further uncertainties in Monte Carlo simulation, such as stochastic variabilities, e.g. process equipment failures (Michael D. McKay, Morrison, & Upton, 1999).

3.3 Sensitivity analysis of the risk quantification

The aim of the sensitivity analysis was to determine the sources of high uncertainty and to quantify their impact on the biomass prediction as the biomass predictions were the basis for the risk quantification. To achieve this, a linear regression was performed on the final biomass predictions (after 6 h) of the Monte Carlo simulation.

The results of the 200 model simulations were a probability distribution (see the histogram of the final yield in Figure 4 A) indicating a considerable variance in the predicted biomass concentration. The variance was therefore decomposed by a linear regression considering the uncertain model parameters, initial conditions, and the ammonia addition that were taken into account in the Monte Carlo procedure. The standardized regression coefficients (SRCs) are presented in Table V. The applied method could satisfactorily decompose the variance with a linear model as the coefficient of determination was $R^2 = 0.96$ for the biomass concentration in this study.

Since $R^2 > 0.7$ the individual impact of each considered parameter on the biomass prediction could be assessed. For example, $\Delta_{NH,add}$, which accounted for the uncertainty in the on-line base addition readout, contributed by 85 % ($SRC^2 = 0.92^2 \cdot 100$ %) to the variance in the biomass prediction (Table V).

In the soft sensor, $\Delta_{NH,}$add ($\Delta_{NH,}$add ~ N(1,0.001)) was multiplied with the on-line measured ammonia addition rate to account for the uncertainty in the base addition. A discussion about the assumed range of uncertainty can be found elsewhere (Spann et al., 2018). Generally speaking, the considered uncertainties of the on-line measurements and the initial conditions are usually evaluated case-by-case incorporating expert inputs from the process engineering and development perspectives.

Table V. Standardized regression coefficient (SRC) of the linear model fitting the biomass concentration prediction with respect to the uncertain parameters.

Group	Parameter name	SRC on biomass production	Sum of SRC^2
Equipment	$\Delta_{NH,add}$	0.92	0.85
Biokinetic model	pH_{opt}	- 0.28	0.11
	$K_{P,La}$	- 0.12	
	σ_{pH}	0.09	
	K_{La}	0.05	
	α	- 0.04	
	Y_{gal}	- 0.04	
Initial conditions	$C_{S,t=0}$	0.04	< 0.01
	$C_{X,t=0}$	- 0.01	

To evaluate the contribution of the parameters on the variance of the biomass prediction, the parameters were grouped to be (i) related to the equipment; (ii) biokinetic parameters; and (iii) initial conditions (Table V). The sum of SRC^2 of the groups was the variance explained by the parameters comprised in the groups. The biomass growth prediction was most sensitive to the uncertainties of the used equipment, namely the balance readout of the base addition ($\Delta_{NH,add}$), as discussed above. The biokinetic model parameters were responsible for 11 % of the total variance of the biomass prediction. The pH function f_{pH} with the parameters pH_{opt} and σ_{pH} (Eqs. (3)-(4)) contributed mainly to it.

Both initial conditions, the initial substrate concentration and the biomass concentration, had a low effect on the final biomass production according to the sensitivity analysis. Substrate was added in excess, and the lactic acid bacteria did not consume everything until the end of the cultivation (see Figure 2 C). This might be attributed to the lactic acid inhibition at the end of cultivation when the lactic acid concentration was high. According to the sensitivity analysis, the initial substrate could therefore be reduced without affecting the final biomass production. The initial biomass concentration had also a minor effect on the final biomass concentration following the results of the sensitivity analysis. However, it would not be possible to reduce the initial biomass concentration without any consequences. A lower initial biomass concentration would lead to a longer lag time and hence a longer cultivation time. The effect of the initial biomass concentration on the lag time was however not captured in the applied sensitivity analysis, as the lag time was updated on-line based on the measured pH. Furthermore, the cultivation time was not considered in the present analysis, instead the yield after 6 h of cultivation was evaluated. Under these conditions, the equipment uncertainty and the pH related parameters were most significant.

Nevertheless, the results of the sensitivity analysis are conditional on the assumptions made in the design of the problem, e.g. which uncertainties were considered as an input and which model application was included (Helton, 1993; Walker et al., 2003). For that reason, the presented results can only be interpreted in the context the problem was designed (Sin et al., 2011).

Within the scope of this study, the sensitivity analysis revealed that $\Delta_{NH,add}$ had the highest effect on the biomass prediction. In other words, the on-line base addition measurement that is a balance readout had a large effect on the quality of the soft sensor prediction. This agrees with the process engineering understanding of the system as the soft sensor uses the base addition to update the μ_{max} parameter. As a result, the risk quantification was also mostly dependent on the base addition measurement because the risk quantification was based on the biomass prediction.

It is consequently essential for an accurate model and risk prediction that the on-line balance measurements are as precise and faulty measurements could be ruled out. This could be for example achieved with a doubled measurement applying a supplemented flow measurement in addition to the balance readout of the base addition.

4 Conclusion

An on-line risk assessment tool was proposed to quantify both pH gradients and the risk of not achieving the target production in a lactic acid bacteria cultivation. To this end, a soft sensor was applied as a PAT tool that was based on a mechanistic model and a CFD-based compartment model. It provided, on the one hand, an on-line prediction of the pH gradient in the bioreactor, which is a critical process parameter. This would enable plant operators to assess the mixing and the base addition strategy. On the other hand, the soft sensor quantified the risk of not achieving the target biomass production. The likelihood of the undesired event, i.e. the target biomass production could not be achieved, was calculated based on the probabilistic model predictions that were obtained from the Monte Carlo simulation of the soft sensor model. The Monte Carlo simulation was performed to consider uncertainties in the model parameters, on-line measurements, and initial process conditions. In the investigated 700 L cultivation, the risk was to lose max. 140 g biomass per batch. The sensitivity analysis revealed that the on-line base addition measurement had the highest impact on the biomass prediction, and hence the soft sensor risk quantification requires accurate on-line measurements. The future objective of this study is the implementation of the soft sensor for risk-based decision making and control in large-scale cultivations under consideration of techno-economic risks.

5 Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 643056 (Biorapid project). We are thankful for the cooperation with Chr. Hansen A/S.

6 Nomenclature

a	Ordinate intercept of the SRC model
b_i	Linear regression coefficient for the i-th model parameter
C_{Gal}	galactose concentration (g L^{-1})
C_{Glc}	glucose concentration (g L^{-1})
C_{LA}	lactate concentration (g L^{-1})
C_{OH^-}	OH^- concentration (mol L^{-1})
C_P	total lactic acid (lactate and lactic acid) concentration (g L^{-1})
C_S	lactose (substrate) concentration (g L^{-1})
C_{tCO}	total carbonic acid ($H_2CO_3^*$ and HCO_3^-) concentration (mol L^{-1})
C_{tNH}	total concentration of NH_4^+ and NH_3 (g L^{-1})
C_{tPh}	total concentration of H_3PO_4, $H_2PO_4^-$, and HPO_4^{2-}(g L^{-1})
C_{tZ}	total concentration of the unknown compound (dissociated and undissociated form) (mol L^{-1})
C_X	biomass concentration (g L^{-1})
f_d	divalent activity coefficients (-)
f_{lag}	lag-time function (-)
f_m	monovalent activity coefficients (-)

f_P	lactic acid inhibition function (-)
f_{pH}	pH dependency function (-)
f_S	substrate limitation and inhibition function (-)
$H_2CO_3^*$	dissolved CO_2 and H_2CO_3
I	ionic strength (g L^{-1})
K'_{C1}	apparent equilibrium constant for the carbonic acid system (-)
K_I	substrate inhibition parameter (g L^{-1})
K_{La}	lactate inhibition parameter (g L^{-1})
K_{La1}	pH dependent lactate inhibition parameter (g L^{-1})
K'_{LA}	apparent equilibrium constant for the lactic acid system (-)
K'_{NH}	apparent equilibrium constant for the ammonia system (-)
K_P	P-controller controller gain
$K_{P,La}$	2. lactate inhibition parameter (L g^{-1})
$K_{P,pH1}$	lactate inhibition pH parameter (-)
$K_{P,pH2}$	2. lactate inhibition pH parameter (-)
K'_{P1}	apparent equilibrium constant for the phosphoric acid system (-)
K'_{P2}	apparent equilibrium constant for the dihydrogen phosphate system (-)
$K'_{r,C1}$	apparent reverse rate constant for carbonic acid dissociation (s^{-1})
$K'_{r,LA}$	apparent reverse rate constant for lactic acid dissociation (s^{-1})
$K'_{r,NH}$	apparent reverse rate constant for NH_4 dissociation (s^{-1})
$K'_{r,P1}$	apparent reverse rate constant for H_3PO_4 dissociation (s^{-1})
$K'_{r,P2}$	apparent reverse rate constant for $H_2PO_4^-$ dissociation (s^{-1})
$K'_{r,W}$	apparent reverse rate constant for water dissociation (s^{-1})
$K'_{r,Z}$	apparent reverse rate constant for the dissociation of the unknown component (s^{-1})
K_S	substrate limitation parameter (g L^{-1})
K'_W	apparent equilibrium constant for the water system (-)
K'_Z	apparent equilibrium constant for the unspecified compound system (-)
n	number of measurements
pH_{opt}	optimal pH parameter in the pH function (-)
pH_{set}	pH control set point (-)
pK_{C1}	pK_a constant for carbonic acid dissociation
pK_{LA}	pK_a constant for lactic acid dissociation
pK_{NH}	pK_a constant for NH_4 dissociation
pK_{P1}	pK_a constant for H_3PO_4 dissociation
pK_{P2}	pK_a constant for $H_2PO_4^-$ dissociation
pK_W	pK_a constant for water dissociation
pK_Z	pK_a constant for the unspecified compound dissociation
q_{Gal}	volumetric galactose secretion rate (C-mol L^{-1} h^{-1})
q_{NH}	volumetric ammonia consumption rate (mol L^{-1} h^{-1})
q_P	volumetric lactic acid secretion rate (C-mol L^{-1} h^{-1})

q_{Ph}	volumetric phosphoric acid consumption rate (mol L^{-1} h^{-1})
q_S	volumetric substrate consumption rate (C-mol L^{-1} h^{-1})
q_X	volumetric biomass growth rate (C-mol L^{-1} h^{-1})

RMSSE	root mean sum of squared errors (g L^{-1})
SRC_i	standardized regression coefficient of the i-th parameter
T	temperature in the bioreactor (K)
t	time variable (h)
t_{lag}	lag-time coefficient (h)
Y_{gal}	galactose yield (g g^{-1})
z_i	charge number of the i-th ion
$\hat{y}_i$	i-th model value of one output (g L^{-1})
$y_{meas,i}$	i-th measurement value of one output (g L^{-1})

Greek Letters

α	growth related production coefficient of lactic acid (g g^{-1})
$\hat{\theta}_{i,j}$	i-th parameter value used on the j-th Monte Carlo simulation
μ_{max}	maximum specific growth rate (h^{-1})
σ	standard deviation
σ_{pH}	spread parameter is the gaussian pH function
$\sigma_{\hat{\theta}_i}$	standard deviation of the estimated parameter
$\sigma_{\hat{y}_X}$	standard deviation of the biomass concentration distribution

7 References

Adam, S., Suzzi, D., Radeke, C., & Khinast, J. G. (2011). An integrated Quality by Design (QbD) approach towards design space definition of a blending unit operation by Discrete Element Method (DEM) simulation. *European Journal of Pharmaceutical Sciences*, *42*(1–2), 106–115. http://doi.org/10.1016/j.ejps.2010.10.013

Aghababaie, M., Khanahmadi, M., & Beheshti, M. (2015). Developing a detailed kinetic model for the production of yogurt starter bacteria in single strain cultures. *Food and Bioproducts Processing*, *94*(April), 657–667. http://doi.org/10.1016/j.fbp.2014.09.007

Åkerberg, C., Hofvendahl, K., Hahn-Hägerdal, B., & Zacchi, G. (1998). Modelling the influence of pH, temperature, glucose and lactic acid concentrations on the kinetics of lactic acid production by Lactococcus lactis ssp. lactis ATCC 19435 in whole-wheat flour. *Applied Microbiology and Biotechnology*, *49*(6), 682–690. http://doi.org/10.1007/s002530051232

Cameron, I., & Raman, R. (2005). *Process Systems Risk Management*. Elsevier Academic Press.

Campolongo, F. (1997). Sensitivity analysis of an environmental model: an application of

different analysis methods. *Reliability Engineering & System Safety*, *57*(1), 49–69. http://doi.org/10.1016/S0951-8320(97)00021-5

Davies, C. W. (1962). *Ion Association*. Londond: Butterworth.

Dawson, R. M. C. (1969). *Data for Biochemical Research*. Oxford: Clarendon Press.

FDA. (2004). *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. U.S. Food and Drug Administration, U.S. Department of Health and Human Services.

Gargalo, C. L., Cheali, P., Posada, J. A., Gernaey, K. V., & Sin, G. (2016). Economic Risk Assessment of Early Stage Designs for Glycerol Valorization in Biorefinery Concepts. *Industrial and Engineering Chemistry Research*, *55*(24), 6801–6814. http://doi.org/10.1021/acs.iecr.5b04593

Glassey, J., Gernaey, K. V., Clemens, C., Schulz, T. W., Oliveira, R., Striedner, G., & Mandenius, C. F. (2011). Process analytical technology (PAT) for biopharmaceuticals. *Biotechnology Journal*, *6*(4), 369–377. http://doi.org/10.1002/biot.201000356

Gnoth, S., Jenzsch, M., Simutis, R., & Lübbert, A. (2007). Process Analytical Technology (PAT): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control. *Journal of Biotechnology*, *132*(2), 180–186. http://doi.org/10.1016/j.jbiotec.2007.03.020

Hasanly, A., Khajeh Talkhoncheh, M., & Karimi Alavijeh, M. (2017). Techno-economic assessment of bioethanol production from wheat straw: a case study of Iran. *Clean Technologies and Environmental Policy*, *20*(2), 357–377. http://doi.org/10.1007/s10098-017-1476-0

Helton, J. C. (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering & System Safety*, *42*(2–3), 327–367. http://doi.org/10.1016/0951-8320(93)90097-I

ICH Q8(R2). (2009). Pharmaceutical Development. Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf

ICH Q9. (2005). Quality Risk Management. Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf

Jiménez-Hornero, J. E., Santos-Dueñas, I. M., & García-García, I. (2009). Optimization of biotechnological processes. The acetic acid fermentation. Part III: Dynamic optimization. *Biochemical Engineering Journal*, *45*(1), 22–29. http://doi.org/10.1016/j.bej.2009.01.011

Kager, J., Herwig, C., & Stelzer, I. V. (2018). State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. *Chemical Engineering Science*, *177*, 234–244. http://doi.org/10.1016/j.ces.2017.11.049

Loewenthal, R. E., Ekama, G. A., & Marais, G. R. (1989). Mixed weak acid/base systems Part I - Mixture characterisation. *Water SA*, *15*(1), 3–24.

Mandenius, C. F., & Gustavsson, R. (2015). Mini-review: Soft sensors as means for PAT in the manufacture of bio-therapeutics. *Journal of Chemical Technology and Biotechnology*, *90*(2), 215–227. http://doi.org/10.1002/jctb.4477

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245. http://doi.org/10.1080/00401706.1979.10489755

McKay, M. D., Morrison, J. D., & Upton, S. C. (1999). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, *117*(1), 44–51. http://doi.org/10.1016/S0010-4655(98)00155-6

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., & Gernaey, K. V. (2017). Mechanistic Fermentation Models for Process Design, Monitoring, and Control. *Trends in Biotechnology*, *35*(10), 914–924. http://doi.org/10.1016/j.tibtech.2017.07.002

Mears, L., Stocks, S. M., Sin, G., & Gernaey, K. V. (2017). A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *Journal of Biotechnology*, *245*, 34–46. http://doi.org/10.1016/j.jbiotec.2017.01.008

Mercier, S. M., Diepenbroek, B., Wijffels, R. H., & Streefland, M. (2014). Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations. *Trends in Biotechnology*, *32*(6), 329–336. http://doi.org/10.1016/j.tibtech.2014.03.008

Musvoto, E. V., Wentzel, M. C., Loewenthal, R. E., & Ekama, G. A. (2000). Integrated Chemical-Physical Processes Modelling - I. Development of a Kinetic-Based Model for Mixed Weak Acid/Base Systems. *Water Research*, *34*(6), 1857–1867. http://doi.org/10.1016/S0043-1354(99)00334-6

Peng, R. Y., Yang, T. C. K., Wang, H., Lin, Y., & Cheng, C. (1997). Modelling of Lactic Acid Fermentation - An Improvement of Leudeking's Model. *Journal of the Chinese Agricultural Chemical Society*, *35*(5), 485–494.

Petrides, D. P., & Siletti, C. A. (2004). The Role of Process Simulation and Scheduling Tools in the Development and Manufacturing of Biopharmaceuticals. In *Proceedings of the 2004 Winter Simulation Conference, 2004.* (Vol. 2, pp. 960–965). IEEE. http://doi.org/10.1109/WSC.2004.1371568

Rantanen, J., & Khinast, J. (2015). The Future of Pharmaceutical Manufacturing Sciences. *Journal of Pharmaceutical Sciences*, *104*(11), 3612–3638. http://doi.org/10.1002/jps.24594

Rathore, A. S., Mittal, S., Pathak, M., & Mahalingam, V. (2014). Chemometrics application in biotech processes: Assessing comparability across processes and scales. *Journal of Chemical Technology and Biotechnology*, *89*(9), 1311–1316. http://doi.org/10.1002/jctb.4428

Rathore, A. S., & Winkle, H. (2009). Quality by design for biopharmaceuticals. *Nature Biotechnology*, *27*(1), 26–34. http://doi.org/10.1038/nbt0109-26

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., … Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley \& Sons, Ltd. http://doi.org/10.1002/9780470725184

Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*, *91*(10–11), 1109–1125. http://doi.org/10.1016/j.ress.2005.11.014

Schepers, A. W., Thibault, J., & Lacroix, C. (2002). Lactobacillus helveticus growth and lactic acid production during pH-controlled batch cultures in whey permeate/yeast extract medium. Part II: kinetic modeling and model validation. *Enzyme and Microbial Technology*, *30*(2), 187–194. http://doi.org/10.1016/S0141-0229(01)00466-5

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M., & Gujer, W. (2009). Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, *43*(11), 2894–2906. http://doi.org/10.1016/j.watres.2009.03.048

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M., & Gujer, W. (2011). Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. *Water Research*, *45*(2), 639–651. http://doi.org/10.1016/j.watres.2010.08.025

Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., … Steckenreiter, T. (2017). Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. *Chemie-Ingenieur-Technik*, *89*(5), 542–561. http://doi.org/10.1002/cite.201600175

Sommeregger, W., Sissolak, B., Kandra, K., von Stosch, M., Mayer, M., & Striedner, G. (2017). Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. *Biotechnology Journal*, *12*(7), 1600546. http://doi.org/10.1002/biot.201600546

Spann, R., Roca, C., Kold, D., Eliasson Lantz, A., Gernaey, K. V., & Sin, G. (2018). A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. *Biochemical Engineering Journal*, *135*, 49–60. http://doi.org/10.1016/j.bej.2018.03.016

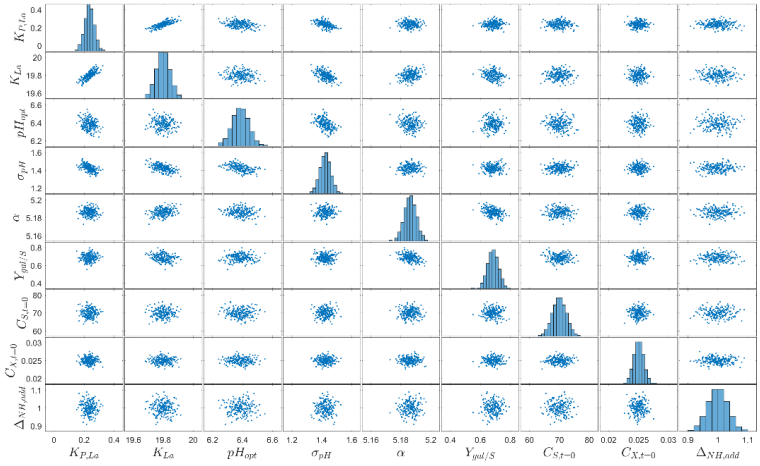
Stocker, E., Toschkoff, G., Sacher, S., & Khinast, J. G. (2014). Use of mechanistic simulations as a quantitative risk-ranking tool within the quality by design framework. *International Journal of Pharmaceutics*, *475*(1), e245–e255. http://doi.org/10.1016/j.ijpharm.2014.08.055

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction Engineering Principles*. Boston, MA: Springer US. http://doi.org/10.1007/978-1-4419-9688-6

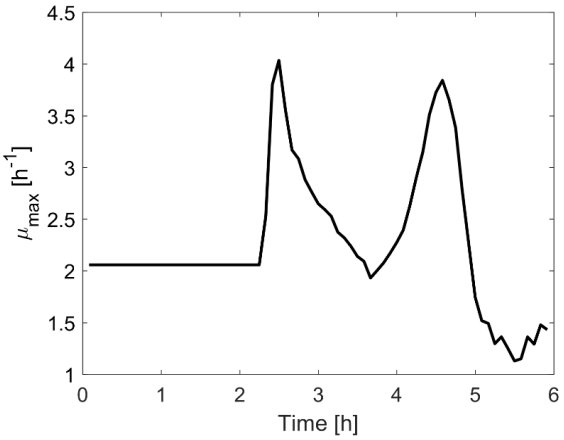
Vrabel, P., Van Der Lans, R. G. J. M., Luyben, K. C. A. M., Boon, L., & Nienow, A. W. (2000). Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: Modelling and measurements. *Chemical Engineering Science*, *55*(23), 5881–5896. http://doi.org/10.1016/S0009-2509(00)00175-5

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment*, *4*(1), 5–17. http://doi.org/10.1076/iaij.4.1.5.16466

8 Supplementary material



Supplementary Figure S1. Plotting of the sampling matrix for the input uncertainty space. The Latin Hypercube Sampling (LHS) technique and the Iman Conover rank correlation method were used to sample 100 independent inputs and to induce the known covariance matrix, respectively.



Supplementary Figure S2. Updated μ_{max} value during the course of the cultivation.

Supplementary calculation of the target biomass production for the risk quantification

The target yield for biomass was 0.09 ± 0.003 g biomass per g lactose based on previous 2 L lab-scale experiments (data not shown). The target biomass concentration was therefore 6.3 ± 0.21 g biomass per liter as the initial substrate concentration was 70 g L^{-1}:

$$\mathbf{0.09}\,\frac{\boldsymbol{g\ biomass}}{\boldsymbol{g\ lactose}}\cdot\mathbf{70}\,\frac{\boldsymbol{g\ lactose}}{\boldsymbol{L}}=\mathbf{6.3}\,\frac{\boldsymbol{g\ biomass}}{\boldsymbol{L}} \tag{A.1}$$

Provided the liquid volume of the cultivation with 700 L, the target biomass production was 4410 ± 0.147 g biomass per batch:

$$\mathbf{6.3}\,\frac{\boldsymbol{g\ biomass}}{\boldsymbol{L}}\cdot\mathbf{700}\,\boldsymbol{L}=\mathbf{4410}\,\boldsymbol{g\ biomass} \tag{A.2}$$

Technical University of Denmark
Department of Chemical and Biochemical Engineering
Process and Systems Engineering Center (PROSYS)
Søltofts Plads Building 229
2800 Kgs. Lyngby
Denmark
Phone: (+45) 45 25 25 25
Email: dtu@dtu.dk
www.dtu.dk